GenCore version 5.1.6

OM protein - protein search, using sw model

Run on: October 22, 2004, 09:06:13 ; Search time 40 Seconds
(without alignments)
580.283 Million cell updates/sec

Title: US-10-089-688-2
Perfect score: 1781
Sequence: 1 MPFPPFGKSHKSPADIVKNLK.....RVKPRTRGIRDLKRPAQQEA 350

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 478139 seqs, 66318000 residues

Total number of hits satisfying chosen parameters: 478139

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_AA:*
1: /cgn2_6/ptodata/1/iaa/5A_COMB.pep:*
2: /cgn2_6/ptodata/1/iaa/5B_COMB.pep:*
3: /cgn2_6/ptodata/1/iaa/6A_COMB.pep:*
4: /cgn2_6/ptodata/1/iaa/6B_COMB.pep:*
5: /cgn2_6/ptodata/1/iaa/PCTUS_COMB.pep:*
6: /cgn2_6/ptodata/1/iaa/backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	1614.5	90.7	341	3	US-09-190-965-3	Sequence 3, Appli
2	1614.5	90.7	341	3	US-09-470-253-3	Sequence 3, Appli
3	1297.5	72.9	337	3	US-09-190-965-1	Sequence 1, Appli
4	1297.5	72.9	337	3	US-09-470-253-1	Sequence 1, Appli
5	1117	62.7	339	3	US-09-190-965-4	Sequence 4, Appli
6	1117	62.7	339	3	US-09-470-253-4	Sequence 4, Appli
7	1054.5	59.2	377	3	US-09-190-965-5	Sequence 5, Appli
8	1054.5	59.2	377	3	US-09-470-253-5	Sequence 5, Appli
9	358.5	20.1	165	4	US-09-248-796A-14303	Sequence 14303, A
10	166	9.3	125	4	US-09-248-796A-15202	Sequence 15202, A
11	163	9.2	90	4	US-09-248-796A-15201	Sequence 15201, A
12	115.5	6.5	630	4	US-09-248-796A-20275	Sequence 20275, A
13	113	6.3	541	4	US-09-134-000C-5420	Sequence 5420, Ap
14	113	6.3	2184	4	US-09-417-485D-6	Sequence 6, Appli
15	110.5	6.2	533	3	US-09-134-001C-4053	Sequence 4053, Ap
16	106	6.0	3959	2	US-08-970-269A-30	Sequence 30, Appl
17	106	6.0	3959	3	US-09-407-562-30	Sequence 30, Appl
18	105.5	5.9	3878	4	US-09-914-259-11	Sequence 11, Appl
19	105	5.9	496	4	US-09-543-681A-6465	Sequence 6465, Ap
20	105	5.9	627	4	US-09-248-796A-20614	Sequence 20614, A
21	105	5.9	776	1	US-08-021-601-2	Sequence 2, Appli
22	105	5.9	776	1	US-08-082-849B-2	Sequence 2, Appli
23	105	5.9	776	5	PCT-US94-01624-2	Sequence 2, Appli
24	104	5.8	933	4	US-09-248-796A-20513	Sequence 20513, A
25	103.5	5.8	1388	3	US-09-572-191-2	Sequence 2, Appli
26	103.5	5.8	1388	3	US-09-723-262-2	Sequence 2, Appli
27	103.5	5.8	1388	3	US-09-723-219-2	Sequence 2, Appli
28	103.5	5.8	2474	4	US-08-265-967C-3	Sequence 3, Appli
29	103.5	5.8	2474	4	US-08-305-790B-4	Sequence 4, Appli
30	103	5.8	1155	4	US-09-543-681A-6286	Sequence 6286, Ap
31	103	5.8	1388	2	US-08-685-576-4	Sequence 4, Appli
32	101.5	5.7	467	4	US-09-543-681A-4262	Sequence 4262, Ap
33	101	5.7	489	4	US-09-248-796A-19030	Sequence 19030, A
34	101	5.7	506	4	US-09-107-532A-5994	Sequence 5994, Ap
35	100.5	5.6	478	4	US-09-328-352-6512	Sequence 6512, Ap
36	100.5	5.6	1388	2	US-08-685-576-1	Sequence 1, Appli
37	100	5.6	460	3	US-08-942-012B-33	Sequence 33, Appl
38	100	5.6	831	4	US-09-269-861A-8	Sequence 8, Appli
39	99	5.6	2954	4	US-09-150-867-1	Sequence 1, Appli
40	98.5	5.5	568	4	US-09-710-279-2482	Sequence 2482, Ap
41	98.5	5.5	967	4	US-09-543-681A-6407	Sequence 6407, Ap
42	98.5	5.5	1145	4	US-09-710-279-1432	Sequence 1432, Ap
43	98.5	5.5	1151	3	US-09-134-001C-3242	Sequence 3242, Ap
44	98	5.5	1388	4	US-09-976-594-296	Sequence 296, App
45	98	5.5	2662	4	US-09-595-684B-31	Sequence 31, Appl

ALIGNMENTS

RESULT 1
US-09-190-965-3
; Sequence 3, Application US/09190965
; Patent No. 6071721
; GENERAL INFORMATION:
; APPLICANT: Tang, Y. Tom
; APPLICANT: Guegler, Karl J.
; APPLICANT: Corley, Neil C.
; APPLICANT: Gorgone, Gina A.
; TITLE OF INVENTION: CALCIUM BINDING PROTEIN
; FILE REFERENCE: PF-0635 US
; CURRENT APPLICATION NUMBER: US/09/190,965
; CURRENT FILING DATE: 1998-11-13
; NUMBER OF SEQ ID NOS: 5
; SOFTWARE: PERL Program
; SEQ ID NO 3
; LENGTH: 341
; TYPE: PRT
; ORGANISM: Mus sp.
; FEATURE: -
; OTHER INFORMATION: g262934
US-09-190-965-3

Query Match 90.7%; Score 1614.5; DB 3; Length 341;
Best Local Similarity 92.9%; Pred. No. 3.1e-151;
Matches 325; Conservative 2; Mismatches 14; Indels 9; Gaps 2;

```
Qy          1 MPFPPFGKSHKSPADIVKNLKESMAVLEKQDISDKKAEKATEEVSKNLVAMKEILYGTNEK 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MPFPPFGKSHKSPADIVKNLKESMAVLEKQDISDKKAEKATEEVSKNLVAMKEILYGTNEK 60

Qy         61 EPQTEAVAQLAQELYNSGLLSTLVADLQLIDFEGKKDVAQIFNNILRRQIGTRTPTVEYI 120
              |||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||||
Db         61 EPQTEAVAQLAQELYNSGLLGTLVADLQLIDFEGKKDVAQIFNNILRRQIGTRTPTVEYI 120

Qy        121 CTQQNILFMLLKGYESPEIALNCGIMLRECIRHEPLAKIILWSEQFYDFFRYVEMSTFDI 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CTQQNILFMLLKGYESPEIALNCGIMLRECIRHEPLAKIILWSEQFYDFFRYVEMSTFDI 180

Qy        181 ASDAFATFKDLLTRHKLLSAEFLEQHYDRFFSEYEKLLHSENYVTKRQSLKLLGELLLDR 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 ASDAFATFKDLLTRHKLLSAEFLEQHYDRFFSEYEKLLHSENYVTKRQSLKLLGELLLDR 240

Qy        241 HNFTIMTKYISKPENLKLMMNLLRDKSRNIQFEAFHVFKVFVANPNKTQPILDILLKNQA 300
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 HNFTIMTKYISKPENLKLMMNLLRDKSRNIQFEAFHVFKVFVANPNKTQPILDILLKNQT 300

Qy        301 KLIEFLSKFQNDRTDCMSSSVPTTNSRVDLRVKPRTRGIRDLKRPAQQEA 350
              |||||||||||||:  |   ||    ||:||| |||||
Db        301 KLIEFLSKFQNDRTE-----DEQFNDEKTYLVKQ----IRNLKRAAQQEA 341
```

```
RESULT 2
US-09-470-253-3
; Sequence 3, Application US/09470253
; Patent No. 6365371
; GENERAL INFORMATION:
;  APPLICANT: Tang, Y. Tom
;  APPLICANT:  Guegler, Karl J.
;  APPLICANT:  Corley, Neil C.
;  APPLICANT:  Gorgone, Gina A.
;  TITLE OF INVENTION: CALCIUM BINDING PROTEIN
;  FILE REFERENCE: PF-0635 US
;  CURRENT APPLICATION NUMBER: US/09/470,253
;  CURRENT FILING DATE:  1999-12-22
;  PRIOR APPLICATION NUMBER: 09/190,965
;  PRIOR FILING DATE: 1998-11-13
;  NUMBER OF SEQ ID NOS: 5
;  SOFTWARE: PERL Program
; SEQ ID NO 3
;   LENGTH: 341
;   TYPE: PRT
;   ORGANISM: Mus sp.
;   FEATURE: -
;   OTHER INFORMATION: g262934
US-09-470-253-3

  Query Match             90.7%;  Score 1614.5;  DB 3;  Length 341;
  Best Local Similarity   92.9%;  Pred. No. 3.1e-151;
  Matches  325;  Conservative    2;  Mismatches   14;  Indels    9;  Gaps    2;

Qy          1 MPFPPFGKSHKSPADIVKNLKESMAVLEKQDISDKKAEKATEEVSKNLVAMKEILYGTNEK 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MPFPPFGKSHKSPADIVKNLKESMAVLEKQDISDKKAEKATEEVSKNLVAMKEILYGTNEK 60

Qy         61 EPQTEAVAQLAQELYNSGLLSTLVADLQLIDFEGKKDVAQIFNNILRRQIGTRTPTVEYI 120
              |||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||||
Db         61 EPQTEAVAQLAQELYNSGLLGTLVADLQLIDFEGKKDVAQIFNNILRRQIGTRTPTVEYI 120

Qy        121 CTQQNILFMLLKGYESPEIALNCGIMLRECIRHEPLAKIILWSEQFYDFFRYVEMSTFDI 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CTQQNILFMLLKGYESPEIALNCGIMLRECIRHEPLAKIILWSEQFYDFFRYVEMSTFDI 180

Qy        181 ASDAFATFKDLLTRHKLLSAEFLEQHYDRFFSEYEKLLHSENYVTKRQSLKLLGELLLDR 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 ASDAFATFKDLLTRHKLLSAEFLEQHYDRFFSEYEKLLHSENYVTKRQSLKLLGELLLDR 240

Qy        241 HNFTIMTKYISKPENLKLMMNLLRDKSRNIQFEAFHVFKVFVANPNKTQPILDILLKNQA 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 HNFTIMTKYISKPENLKLMMNLLRDKSRNIQFEAFHVFKVFVANPNKTQPILDILLKNQT 300

Qy        301 KLIEFLSKFQNDRTDCMSSSVPTTNSRVDLRVKPRTRGIRDLKRPAQQEA 350
              |||||||||||||:         |      ||     ||:||| |||||
Db        301 KLIEFLSKFQNDRTE-----DEQFNDEKTYLVKQ----IRNLKRAAQQEA 341


RESULT 3
US-09-190-965-1
; Sequence 1, Application US/09190965
; Patent No. 6071721
; GENERAL INFORMATION:
;  APPLICANT: Tang, Y. Tom
;  APPLICANT: Guegler, Karl J.
;  APPLICANT: Corley, Neil C.
;  APPLICANT: Gorgone, Gina A.
;  TITLE OF INVENTION: CALCIUM BINDING PROTEIN
;  FILE REFERENCE: PF-0635 US
;  CURRENT APPLICATION NUMBER: US/09/190,965
;  CURRENT FILING DATE: 1998-11-13
;  NUMBER OF SEQ ID NOS: 5
;  SOFTWARE: PERL Program
; SEQ ID NO 1
;   LENGTH: 337
;   TYPE: PRT
;   ORGANISM: Homo sapiens
;   FEATURE: -
;   OTHER INFORMATION: 3734805
US-09-190-965-1

  Query Match             72.9%;  Score 1297.5;  DB 3;  Length 337;
  Best Local Similarity   76.0%;  Pred. No. 6.6e-120;
  Matches  263;  Conservative   32;  Mismatches   38;  Indels   13;  Gaps    3;

Qy          1 MPFPPFGKSHKSPADIVKNLKESMAVLEKQDISDKKAEKATEEVSKNLVAMKEILYGTNEK 60
              ||  | ||||:||:||| ||:::|:||||   ||| :||:|||||:| ||||||  |||||
Db          4 MPL-FSKSHKNPAEIVKILKDNLAILEKQ---DKKTDKASEEVSKSLQAMKEILCGTNEK 59

Qy         61 EPQTEAVAQLAQELYNSGLLSTLVADLQLIDFEGKKDVAQIFNNILRRQIGTRTPTVEYI 120
              || ||||||||||||:|||| ||:|||||||||||||| |||||||||||||:||||||
Db         60 EPPTEAVAQLAQELYSSGLLVTLIADLQLIDFEGKKDVTQIFNNILRRQIGTRSPTVEYI 119

Qy        121 CTQQNILFMLLKGYESPEIALNCGIMLRECIRHEPLAKIILWSEQFYDFFRYVEMSTFDI 180
                :||||||||||:|:||| ||||||||||||||||||:| || |||:|||:|||||
Db        120 SAHPHILFMLLKGYEAPQIALRCGIMLRECIRHEPLAKIILFSNQFRDFFKYVELSTFDI 179

Qy        181 ASDAFATFKDLLTRHKLLSAEFLEQHYDRFFSEYEKLLHSENYVTKRQSLKLLGELLLDR 240
              ||||||||||||||||:| |:||||:||  | :||||| ||||||||||||||||:|||
Db        180 ASDAFATFKDLLTRHKVLVADFLEQNYDTIFEDYEKLLQSENYVTKRQSLKLLGELILDR 239

Qy        241 HNFTIMTKYISKPENLKLMMNLLRDKSRNIQFEAFHVFKVFVANPNKTQPILDILLKNQA 300
              ||| ||||||||||||||||||||| |||||||||||||||:|:|||||::||||||
Db        240 HNFAIMTKYISKPENLKLMMNLLRDKSPNIQFEAFHVFKVFVASPHKTQPIVEILLKNQP 299

Qy        301 KLIEFLSKFQNDRTDCMSSSVPTTNSRVDLRVKPRTRGIRDLKRPA 346
              ||||||| || :|||       : :       : |||||: |
Db        300 KLIEFLSSFQKERTD---------DEQFADEKNYLIKQIRDLKKTA 336


RESULT 4
US-09-470-253-1
; Sequence 1, Application US/09470253
; Patent No. 6365371
; GENERAL INFORMATION:
;  APPLICANT: Tang, Y. Tom
;  APPLICANT:  Guegler, Karl J.
;  APPLICANT:  Corley, Neil C.
;  APPLICANT:  Gorgone, Gina A.
;  TITLE OF INVENTION: CALCIUM BINDING PROTEIN
;  FILE REFERENCE: PF-0635 US
;  CURRENT APPLICATION NUMBER: US/09/470,253
;  CURRENT FILING DATE:  1999-12-22
;  PRIOR APPLICATION NUMBER: 09/190,965
;  PRIOR FILING DATE: 1998-11-13
;  NUMBER OF SEQ ID NOS: 5
;  SOFTWARE: PERL Program
; SEQ ID NO 1
;   LENGTH: 337
;   TYPE: PRT
;   ORGANISM: Homo sapiens
;   FEATURE: -
;   OTHER INFORMATION: 3734805
US-09-470-253-1

  Query Match             72.9%;  Score 1297.5;  DB 3;  Length 337;
  Best Local Similarity   76.0%;  Pred. No. 6.6e-120;
  Matches  263;  Conservative   32;  Mismatches   38;  Indels   13;  Gaps    3;

Qy          1 MPFPPFGKSHKSPADIVKNLKESMAVLEKQDISDKKAEKATEEVSKNLVAMKEILYGTNEK 60
              ||  | ||||:||:||| ||:::|:||||   ||| :||:|||||:| ||||||  |||||
Db          4 MPL-FSKSHKNPAEIVKILKDNLAILEKQ---DKKTDKASEEVSKSLQAMKEILCGTNEK 59

Qy         61 EPQTEAVAQLAQELYNSGLLSTLVADLQLIDFEGKKDVAQIFNNILRRQIGTRTPTVEYI 120
              || ||||||||||||:|||| ||:|||||||||||||| |||||||||||||:||||||
Db         60 EPPTEAVAQLAQELYSSGLLVTLIADLQLIDFEGKKDVTQIFNNILRRQIGTRSPTVEYI 119

Qy        121 CTQQNILFMLLKGYESPEIALNCGIMLRECIRHEPLAKIILWSEQFYDFFRYVEMSTFDI 180
```

```
Db        120 SAHPHILFMLLKGYEAPQIALRCGIMLRECIRHEPLAKIILFSNQFRDFFKYVELSTFDI 179

Qy        181 ASDAFATFKDLLTRHKLLSAEFLEQHYDRFFSEYEKLLHSENYVTKRQSLKLLGELLLDR 240

Db        180 ASDAFATFKDLLTRHKVLVADFLEQNYDTIFEDYEKLLQSENYVTKRQSLKLLGELILDR 239

Qy        241 HNFTIMTKYISKPENLKLMMNLLRDKSRNIQFEAFHVFKVFVANPNKTQPILDILLKNQA 300

Db        240 HNFAIMTKYISKPENLKLMMNLLRDKSPNIQFEAFHVFKVFVASPHKTQPIVEILLKNQP 299

Qy        301 KLIEFLSKFQNDRTDCMSSSVPTTNSRVDLRVKPRTRGIRDLKRPA 346

Db        300 KLIEFLSSFQKERTD---------DEQFADEKNYLIKQIRDLKKTA 336


RESULT 5
US-09-190-965-4
; Sequence 4, Application US/09190965
; Patent No. 6071721
; GENERAL INFORMATION:
;  APPLICANT: Tang, Y. Tom
;  APPLICANT: Guegler, Karl J.
;  APPLICANT: Corley, Neil C.
;  APPLICANT: Gorgone, Gina A.
;  TITLE OF INVENTION: CALCIUM BINDING PROTEIN
;  FILE REFERENCE: PF-0635 US
;  CURRENT APPLICATION NUMBER: US/09/190,965
;  CURRENT FILING DATE: 1998-11-13
;  NUMBER OF SEQ ID NOS: 5
;  SOFTWARE: PERL Program
; SEQ ID NO 4
;   LENGTH: 339
;   TYPE: PRT
;   ORGANISM: Drosophila melanogaster
;   FEATURE: -
;   OTHER INFORMATION: g1794137
US-09-190-965-4

  Query Match             62.7%;  Score 1117;  DB 3;  Length 339;
  Best Local Similarity   69.0%;  Pred. No. 4.7e-102;
  Matches  220;  Conservative   51;  Mismatches   40;  Indels    8;  Gaps    5;

Qy          1 MPFPFGKSHKSPADIVKNLKESMAVLEKQDISDKKAEKATEEVSKNLVAMKEILYGTNEK 60

Db          1 MPL-FGKSQKSPVELVKSLKEAINALE---AGDRKVEKAQEDVSKNLVSIKNMLHGSSDA 56

Qy         61 EPQTE-AVAQLAQELYNSGLLSTLVADLQLIDFEGKKDVAQIFNNILRRQIGTRTPTVEY 119

Db         57 EPPADYVVAQLSQELYNSNLLLLLIQNLHRIDFEGKKHVALIFNNLLRRQIGTRSPTVEY 116

Qy        120 ICTQQNILFMLLKGYES--PEIALNCGIMLRECIRHEPLAKIILWSEQFYDFFRYVEMST 177

Db        117 ICTKPEILFTLMAGYEDAHPEIALNSGTMLRECARYEALAKIMLHSDEFFKFFRYVEVST 176

Qy        178 FDIASDAFATFKDLLTRHKLLSAEFLEQHYDRFFSE-YEKLLHSENYVTKRQSLKLLGEL 236

Db        177 FDIASDAFSTFKELLTRHKLLCAEFLDANYDKFFSQHYQRLLNSENYVTRRQSLKLLGEL 236

Qy        237 LLDRHNFTIMTKYISKPENLKLMMNLLRDKSRNIQFEAFHVFKVFVANPNKTQPILDILL 296

Db        237 LLDRHNFTVMTRYISEPENLKLMMNMLKEKSRNIQFEAFHVFKVFVANPNKPKPILDILL 296

Qy        297 KNQAKLIEFLSKFQNDRTD 315

Db        297 RNQTKLVDFLTNFHTDRSE 315


RESULT 6
US-09-470-253-4
; Sequence 4, Application US/09470253
; Patent No. 6365371
; GENERAL INFORMATION:
;  APPLICANT: Tang, Y. Tom
;  APPLICANT:  Guegler, Karl J.
;  APPLICANT:  Corley, Neil C.
;  APPLICANT:  Gorgone, Gina A.
;  TITLE OF INVENTION: CALCIUM BINDING PROTEIN
;  FILE REFERENCE: PF-0635 US
;  CURRENT APPLICATION NUMBER: US/09/470,253
;  CURRENT FILING DATE:  1999-12-22
;  PRIOR APPLICATION NUMBER: 09/190,965
;  PRIOR FILING DATE: 1998-11-13
;  NUMBER OF SEQ ID NOS: 5
;  SOFTWARE: PERL Program
; SEQ ID NO 4
;   LENGTH: 339
;   TYPE: PRT
;   ORGANISM: Drosophila melanogaster
;   FEATURE: -
;   OTHER INFORMATION: g1794137
US-09-470-253-4

  Query Match             62.7%;  Score 1117;  DB 3;  Length 339;
  Best Local Similarity   69.0%;  Pred. No. 4.7e-102;
  Matches  220;  Conservative   51;  Mismatches   40;  Indels    8;  Gaps    5;

Qy          1 MPFPFGKSHKSPADIVKNLKESMAVLEKQDISDKKAEKATEEVSKNLVAMKEILYGTNEK 60

Db          1 MPL-FGKSQKSPVELVKSLKEAINALE---AGDRKVEKAQEDVSKNLVSIKNMLHGSSDA 56

Qy         61 EPQTE-AVAQLAQELYNSGLLSTLVADLQLIDFEGKKDVAQIFNNILRRQIGTRTPTVEY 119

Db         57 EPPADYVVAQLSQELYNSNLLLLLIQNLHRIDFEGKKHVALIFNNLLRRQIGTRSPTVEY 116

Qy        120 ICTQQNILFMLLKGYES--PEIALNCGIMLRECIRHEPLAKIILWSEQFYDFFRYVEMST 177

Db        117 ICTKPEILFTLMAGYEDAHPEIALNSGTMLRECARYEALAKIMLHSDEFFKFFRYVEVST 176

Qy        178 FDIASDAFATFKDLLTRHKLLSAEFLEQHYDRFFSE-YEKLLHSENYVTKRQSLKLLGEL 236

Db        177 FDIASDAFSTFKELLTRHKLLCAEFLDANYDKFFSQHYQRLLNSENYVTRRQSLKLLGEL 236

Qy        237 LLDRHNFTIMTKYISKPENLKLMMNLLRDKSRNIQFEAFHVFKVFVANPNKTQPILDILL 296

Db        237 LLDRHNFTVMTRYISEPENLKLMMNMLKEKSRNIQFEAFHVFKVFVANPNKPKPILDILL 296

Qy        297 KNQAKLIEFLSKFQNDRTD 315

Db        297 RNQTKLVDFLTNFHTDRSE 315


RESULT 7
US-09-190-965-5
; Sequence 5, Application US/09190965
; Patent No. 6071721
; GENERAL INFORMATION:
;  APPLICANT: Tang, Y. Tom
;  APPLICANT: Guegler, Karl J.
;  APPLICANT: Corley, Neil C.
;  APPLICANT: Gorgone, Gina A.
;  TITLE OF INVENTION: CALCIUM BINDING PROTEIN
;  FILE REFERENCE: PF-0635 US
;  CURRENT APPLICATION NUMBER: US/09/190,965
;  CURRENT FILING DATE: 1998-11-13
;  NUMBER OF SEQ ID NOS: 5
;  SOFTWARE: PERL Program
; SEQ ID NO 5
;   LENGTH: 377
;   TYPE: PRT
;   ORGANISM: Caenorhabditis elegans
;   FEATURE: -
;   OTHER INFORMATION: g1255838
US-09-190-965-5

  Query Match             59.2%;  Score 1054.5;  DB 3;  Length 377;
```

```
  Best Local Similarity   62.5%;  Pred. No. 8.2e-96;
  Matches  205;  Conservative   49;  Mismatches   61;  Indels   13;  Gaps    2;

Qy          1 MPFPPFGKSHKSPADIVKNLKESMAVLEK----------QDISDKKAEKATEEVSKNLVAM 50
              ||  ||||||||||:|| |:| : :|:|              |||| :|| :|||||:  :
Db          1 MPLLPFGKSHKSPADVVKTLREVLTILDKLPPPKLDKDGNIQSDKKYDKALDEVSKNVAMI 60

Qy         51 KEILYGTNEKEPQTE---AVAQLAQELYNSGLLSTLVADLQLIDFEGKKDVAQIFNNILR 107
              |  :|| :  || :|    |||||||:||: :|  |:  |   :|| |||| |||||:||
Db         61 KSFIYGNDSAEPSSEHVVQVAQLAQEVYNANILPMLIKMLPKFEFECKKDVGQIFNNLLR 120

Qy        108 RQIGTRTPTVEYICTQQNILFMLLKGYESPEIALNCGIMLRECIRHEPLAKIILWSEQFY 167
              ||||||:|||||:  :  ||  |::||  |:||| ||:|||| |||: ||||||:|: ||
Db        121 RQIGTRSPTVEYLGARPEILIQLVQGYSVPDIALTCGLMLRESIRHDHLAKIILYSDVFY 180

Qy        168 DFFRYVEMSTFDIASDAFATFKDLLTRHKLLSAEFLEQHYDRFFSEYEKLLHSENYVTKR 227
              || ||:   |||:||||:|||:| |||| : ||||: :|| ||::|: ||:|:||||:|
Db        181 TFFLYVQSEVFDISSDAFSTFKELTTRHKAIIAEFLDSNYDTFFAQYQNLLNSKNYVTRR 240

Qy        228 QSLKLLGELLLDRHNFTIMTKYISKPENLKLMMNLLRDKSRNIQFEAFHVFKVFVANPNK 287
              |||||||||||||||  |||||| |:||:||| ||||||||||:|||||||||||||||
Db        241 QSLKLLGELLLDRHNFNTMTKYISNPDNLRLMMELLRDKSRNIQYEAFHVFKVFVANPNK 300

Qy        288 TQPILDILLKNQAKLIEFLSKFQNDRTD 315
              :|| ||| :|: ||:||||:| |||||
Db        301 PKPISDILNRNREKLVEFLSEFHNDRTD 328


RESULT 8
US-09-470-253-5
; Sequence 5, Application US/09470253
; Patent No. 6365371
; GENERAL INFORMATION:
;  APPLICANT: Tang, Y. Tom
;  APPLICANT:  Guegler, Karl J.
;  APPLICANT:  Corley, Neil C.
;  APPLICANT:  Gorgone, Gina A.
;  TITLE OF INVENTION: CALCIUM BINDING PROTEIN
;  FILE REFERENCE: PF-0635 US
;  CURRENT APPLICATION NUMBER: US/09/470,253
;  CURRENT FILING DATE:  1999-12-22
;  PRIOR APPLICATION NUMBER: 09/190,965
;  PRIOR FILING DATE: 1998-11-13
;  NUMBER OF SEQ ID NOS: 5
;  SOFTWARE: PERL Program
; SEQ ID NO 5
;   LENGTH: 377
;   TYPE: PRT
;   ORGANISM: Caenorhabditis elegans
;   FEATURE: -
;   OTHER INFORMATION: g1255838
US-09-470-253-5

  Query Match             59.2%;  Score 1054.5;  DB 3;  Length 377;
  Best Local Similarity   62.5%;  Pred. No. 8.2e-96;
  Matches  205;  Conservative   49;  Mismatches   61;  Indels   13;  Gaps    2;

Qy          1 MPFPPFGKSHKSPADIVKNLKESMAVLEK----------QDISDKKAEKATEEVSKNLVAM 50
              ||  ||||||||||:|| |:| : :|:|              |||| :|| :|||||:  :
Db          1 MPLLPFGKSHKSPADVVKTLREVLTILDKLPPPKLDKDGNIQSDKKYDKALDEVSKNVAMI 60

Qy         51 KEILYGTNEKEPQTE---AVAQLAQELYNSGLLSTLVADLQLIDFEGKKDVAQIFNNILR 107
              |  :|| :  || :|    |||||||:||: :|  |:  |   :|| |||| |||||:||
Db         61 KSFIYGNDSAEPSSEHVVQVAQLAQEVYNANILPMLIKMLPKFEFECKKDVGQIFNNLLR 120

Qy        108 RQIGTRTPTVEYICTQQNILFMLLKGYESPEIALNCGIMLRECIRHEPLAKIILWSEQFY 167
              ||||||:|||||:  :  ||  |::||  |:||| ||:|||| |||: ||||||:|: ||
Db        121 RQIGTRSPTVEYLGARPEILIQLVQGYSVPDIALTCGLMLRESIRHDHLAKIILYSDVFY 180

Qy        168 DFFRYVEMSTFDIASDAFATFKDLLTRHKLLSAEFLEQHYDRFFSEYEKLLHSENYVTKR 227
              || ||:   |||:||||:|||:| |||| : ||||: :|| ||::|: ||:|:||||:|
Db        181 TFFLYVQSEVFDISSDAFSTFKELTTRHKAIIAEFLDSNYDTFFAQYQNLLNSKNYVTRR 240

Qy        228 QSLKLLGELLLDRHNFTIMTKYISKPENLKLMMNLLRDKSRNIQFEAFHVFKVFVANPNK 287
              |||||||||||||||  |||||| |:||:||| ||||||||||:|||||||||||||||
Db        241 QSLKLLGELLLDRHNFNTMTKYISNPDNLRLMMELLRDKSRNIQYEAFHVFKVFVANPNK 300

Qy        288 TQPILDILLKNQAKLIEFLSKFQNDRTD 315
              :|| ||| :|: ||:||||:| |||||
Db        301 PKPISDILNRNREKLVEFLSEFHNDRTD 328


RESULT 9
US-09-248-796A-14303
; Sequence 14303, Application US/09248796A
; Patent No. 6747137
; GENERAL INFORMATION:
;  APPLICANT: Keith Weinstock et al
;  TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO CANDIDA ALBICAN
;  TITLE OF INVENTION:  FOR DIAGNOSTICS AND THERAPEUTICS
;  FILE REFERENCE: 107196.132
;  CURRENT APPLICATION NUMBER: US/09/248,796A
;  CURRENT FILING DATE:  1999-02-12
;  PRIOR APPLICATION NUMBER: US 60/074,725
;  PRIOR FILING DATE: 1998-02-13
;  PRIOR APPLICATION NUMBER: US 60/096,409
;  PRIOR FILING DATE: 1998-08-13
;  NUMBER OF SEQ ID NOS: 28208
; SEQ ID NO 14303
;   LENGTH: 165
;   TYPE: PRT
;   ORGANISM: Candida albicans
US-09-248-796A-14303

  Query Match             20.1%;  Score 358.5;  DB 4;  Length 165;
  Best Local Similarity   48.6%;  Pred. No. 1.5e-27;
  Matches   72;  Conservative   32;  Mismatches   43;  Indels    1;  Gaps    1;

Qy         96 KDVAQIFNNILRRQIGTRTPTVEYIC-TQQNILFMLLKGYESPEIALNCGIMLRECIRHE 154
              |||  :|  :||| :  ::|||:|:  ::   |: ||:|| |: || | || :||:||: |
Db         13 KDVVILFLTLLRRTMANKSPTVDYLVHSKPEIITMLIKGPENLEIGLICGQILRDCIKFE 72

Qy        155 PLAKIILWSEQFYDFFRYVEMSTFDIASDAFATFKDLLTRHKLLSAEFLEQHYDRFFSEY 214
              : | :|:|  ||:||:||:: |||||:||  |  :||| |: | :|||  :|| | :
Db         73 VINKFVLYSPSFYNFFKYVQIPTFDIATDAMMTLHELLTTHRKLVSEFLGNNYDVFITAI 132

Qy        215 EKLLHSENYVTKRQSLKLLGELLLDRHN 242
              ||: |:|||||| |:||| ||:  | |
Db        133 NKLVTSKNYVTKRLSVKLLDELVSQRSN 160


RESULT 10
US-09-248-796A-15202
; Sequence 15202, Application US/09248796A
; Patent No. 6747137
; GENERAL INFORMATION:
;  APPLICANT: Keith Weinstock et al
;  TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO CANDIDA ALBICAN
;  TITLE OF INVENTION:  FOR DIAGNOSTICS AND THERAPEUTICS
;  FILE REFERENCE: 107196.132
;  CURRENT APPLICATION NUMBER: US/09/248,796A
;  CURRENT FILING DATE:  1999-02-12
;  PRIOR APPLICATION NUMBER: US 60/074,725
;  PRIOR FILING DATE: 1998-02-13
;  PRIOR APPLICATION NUMBER: US 60/096,409
;  PRIOR FILING DATE: 1998-08-13
;  NUMBER OF SEQ ID NOS: 28208
; SEQ ID NO 15202
;   LENGTH: 125
;   TYPE: PRT
;   ORGANISM: Candida albicans
US-09-248-796A-15202

  Query Match              9.3%;  Score 166;  DB 4;  Length 125;
  Best Local Similarity   34.2%;  Pred. No. 1.1e-08;
  Matches   38;  Conservative   28;  Mismatches   43;  Indels    2;  Gaps    1;
```

```
Qy          1 MPFPPFGKSHKSPADIVKNLKESMAVLEKQDISDKKAEKATEEVSKNLVAMKEILYGTNEK 60
              | | | :: |:| ::|: | :   ||:    |   |:|  :| :: |  || ||:| :|
Db         16 MAFLFKRNPKTPPELVRALNDQ--VLKLDYASPDNAKKYQDECARYLKNMKVILHGDDEV 73

Qy         61 EPQTEAVAQLAQELYNSGLLSTLVADLQLIDFEGKKDVAQIFNNILRRQIG 111
              ||| : : |||||:|::  |  || :|: :||: ::    | ::|:    |
Db         74 EPQPDQITQLAQEIYSTDCLYYLVVNLRKLDFDSRERCCYIVSDIIATNYG 124


RESULT 11
US-09-248-796A-15201
; Sequence 15201, Application US/09248796A
; Patent No. 6747137
; GENERAL INFORMATION:
;  APPLICANT: Keith Weinstock et al
;  TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO CANDIDA ALBICAN
;  TITLE OF INVENTION:  FOR DIAGNOSTICS AND THERAPEUTICS
;  FILE REFERENCE: 107196.132
;  CURRENT APPLICATION NUMBER: US/09/248,796A
;  CURRENT FILING DATE:  1999-02-12
;  PRIOR APPLICATION NUMBER: US 60/074,725
;  PRIOR FILING DATE: 1998-02-13
;  PRIOR APPLICATION NUMBER: US 60/096,409
;  PRIOR FILING DATE: 1998-08-13
;  NUMBER OF SEQ ID NOS: 28208
; SEQ ID NO 15201
;   LENGTH: 90
;   TYPE: PRT
;   ORGANISM: Candida albicans
US-09-248-796A-15201

  Query Match             9.2%;  Score 163;  DB 4;  Length 90;
  Best Local Similarity   50.8%;  Pred. No. 1.3e-08;
  Matches   31;  Conservative    8;  Mismatches   22;  Indels    0;  Gaps    0;

Qy        249 YISKPENLKLMMNLLRDKSRNIQFEAFHVFKVFVANPNKTQPILDILLKNQAKLIEFLSK 308
              :    |||| | || || :|:| | ||  | ||||| ::| : |||:||:|  |||
Db          1 FFDDANNLKLTMLLLSDKLKNLQLEGFHTLKFFVANPKRSQKVTDILIKNKANFIEFFKT 60

Qy        309 F 309
              |
Db         61 F 61


RESULT 12
US-09-248-796A-20275
; Sequence 20275, Application US/09248796A
; Patent No. 6747137
; GENERAL INFORMATION:
;  APPLICANT: Keith Weinstock et al
;  TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO CANDIDA ALBICAN
;  TITLE OF INVENTION:  FOR DIAGNOSTICS AND THERAPEUTICS
;  FILE REFERENCE: 107196.132
;  CURRENT APPLICATION NUMBER: US/09/248,796A
;  CURRENT FILING DATE:  1999-02-12
;  PRIOR APPLICATION NUMBER: US 60/074,725
;  PRIOR FILING DATE: 1998-02-13
;  PRIOR APPLICATION NUMBER: US 60/096,409
;  PRIOR FILING DATE: 1998-08-13
;  NUMBER OF SEQ ID NOS: 28208
; SEQ ID NO 20275
;   LENGTH: 630
;   TYPE: PRT
;   ORGANISM: Candida albicans
US-09-248-796A-20275

  Query Match             6.5%;  Score 115.5;  DB 4;  Length 630;
  Best Local Similarity   20.5%;  Pred. No. 0.012;
  Matches   74;  Conservative   80;  Mismatches  138;  Indels   69;  Gaps   15;

Qy         10 KSPADIVK-NLKESMAVLEKQDISDKKAEKATEEVSK---NLVAMKEILYGTNEKEPQTE 65
              |:  | :| : ||   :  |   :: | ||  ||: |   :|    | | |  |:|

Db        141 KTSGDALKQSQKEYKTLKTKNSDTELKLEKQLEELEKVKLDLQTADEKLKGITERE---I 197

Qy         66 AVAQLAQELYNSGLLSTLVADLQLIDFEGKKDVAQIFNNILRRQIGTRTPTVEYICTQQN 125
              |:    : : |||| :|                                ::   | ||: :  ::
Db        198 ALKSELETVKNSGLSTT------------------------SELAALTKTVKSLEKEKE 232

Qy        126 ILFMLLKGYESPEIALNCGIMLRECI-RHEPLAKIIL--------WSEQFYDFFRYVEMS 176
              |   | | :| |         | : | :|  ::: :           ::|| |     :
Db        233 EL-QFLSGNKSKE--------LEDYIQKHSDISEKLKALTDELKEKTKQFDD----SKKK 279

Qy        177 TFDIASDAFATFKDLLTRHKLLSAEF--LEQHYDRFFSEYEK---LLHSENYVTKRQSLK 231
                :: :|  :| |:| |  |  :::|  ||:  |:   :  |   || ::|   |:: |:
Db        280 LTELENDLTSTKKELET-EKTQTSKFKNLEERKDKEIVKLNKELELLKNDNSGAKKELLE 338

Qy        232 LLGELLLDRHNFTIMTKYISKPENLKLMMNLLRDKSRNIQFEAFHVFKVFVANPNKTQPI 291
              : :|      |::| :   :::    : |:::::    |   | | :    : |:
Db        339 KVSKL---ESEIEILSKELEDKKSVMKQHDELKEQTKEKNQELQKVTKDY----STTKLK 391

Qy        292 LDILLKNQAKLIEFLSKFQ--NDRTDCMSSSVPTTNSRVDLRVKPRTRGIRDLKRPAQQE 349
              || | |     : |  ||:  : :    :| :   | :::: :  : :  ::|::  :|
Db        392 LDELQKELDAALSFKDKFETASAKLTQSTSDLEAANKKLNILISEKEKTEQELEKLTKQH 451

Qy        350 A 350
              |
Db        452 A 452


RESULT 13
US-09-134-000C-5420
; Sequence 5420, Application US/09134000C
; Patent No. 6617156
; GENERAL INFORMATION:
;  APPLICANT: Lynn Doucette-Stamm et al
;  TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO
;  TITLE OF INVENTION:  ENTEROCOCCUS FAECALIS FOR DIAGNOSTICS AND THERAPEUTICS
;  FILE REFERENCE: 032796-032
;  CURRENT APPLICATION NUMBER: US/09/134,000C
;  CURRENT FILING DATE:  1998-08-13
;  PRIOR APPLICATION NUMBER: US 60/055,778
;  PRIOR FILING DATE: 1997-08-15
;  NUMBER OF SEQ ID NOS: 6812
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 5420
;   LENGTH: 541
;   TYPE: PRT
;   ORGANISM: Enterococcus faecalis
US-09-134-000C-5420

  Query Match             6.3%;  Score 113;  DB 4;  Length 541;
  Best Local Similarity   23.0%;  Pred. No. 0.017;
  Matches   89;  Conservative   61;  Mismatches  141;  Indels   96;  Gaps   21;

Qy         14 DIVKNLKESMAVLEKQ-DISD----------KKAEKATEEVSKNLVAMKEILYGTNEKEP 62
              ||:| |  :: :|:|  :|||           : :    | :||  ::  :    | |:
Db        173 DIIK-LPFALKMLDKHLEISDGDLPHDAIVLSREKLKNSEFTKNDYSLLTLYSHANPKK- 230

Qy         63 QTEAVAQLAQE--LYNSGLLSTL---------VADLQLIDFEGKKDVAQ-IFN--NILRR 108
                  ||| |  ::||   : |         | :  |:  |   || |  ||  |:|::
Db        231 ------QLALEVSIFNSYSTNCLLLVYLQDEKVKETDLVIIENAIDVLQEKFNTENLLKK 284

Qy        109 QIGTRTPTVEYICTQQNILFMLLKGYESPEIALNCGIMLRECIRHEPLAKIILWSEQFYD 168
              :  ||          |:   :|:   :|:       :|||    |             |
Db        285 ERYTRL---------NNLADAILQ--NTPQNPDELNSLLREVQMQE-----------LD 321

Qy        169 FFRYVEMSTFDIASDAFATFKDLLTRHKLLSAEFLEQ-HYDRFFSEYEKLLHSENYVTKR 227
               :: :  || |  :          | |   : : | :| :|       : :   |:  :||
Db        322 SYQAIAFSTKDTNTQLMKERIIALLRTLRVKSIFFDQLNYSAVLFNFNE---SDGKITKL 378

Qy        228 QSLKLLGELLLDRHNFTIMTKYISKPENLKLM----MNLLRDKSRNIQFEAFH------- 276
              |  :|| ||| :    |:    :   | :| :    :::||        | |:
Db        379 QLSRLLAELLEENDTLTVAVSSLKSREGIKELLIECLDILRFN------ETFYNGPIVTL 432

Qy        277 ----VFKVFV-----ANPNKTQP-ILDILLKNQAKLIEFL-SKFQNDRTDCMSSSVPTTN 325
```

```
Db        433 ADIGVFKNFIREDQLENLDELIPRALYQLAENNYDLFETLYSFFQNNRNYKQTSEAMFLH 492

Qy        326 SRVDLRVKPRTRGIR-----DLKRPAQ 347
Db        493 SKT---IRYRLNKVEQLLDIDLANPLQ 516


RESULT 14
US-09-417-485D-6
; Sequence 6, Application US/09417485D
; Patent No. 6541202
; GENERAL INFORMATION:
;  APPLICANT: Long, David M.
;  APPLICANT:  Metz, Anneke M.
;  APPLICANT:  Love, Ruschelle A.
;  TITLE OF INVENTION: Telomerase Reverse Transcriptase (TERT) Genes
;  FILE REFERENCE: 47714-5009-US
;  CURRENT APPLICATION NUMBER: US/09/417,485D
;  CURRENT FILING DATE:  2002-06-14
;  NUMBER OF SEQ ID NOS: 49
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 6
;   LENGTH: 2184
;   TYPE: PRT
;   ORGANISM: Plasmodium falciparum
;   FEATURE:
;   NAME/KEY: unsure
;   LOCATION: (330)..(335)
;   OTHER INFORMATION: Xaa at position 330 = Leu or Ile;
;   OTHER INFORMATION: Xaa at position 335 = Asp or Gly.
US-09-417-485D-6

  Query Match             6.3%;  Score 113;  DB 4;  Length 2184;
  Best Local Similarity   22.8%;  Pred. No. 0.13;
  Matches   68;  Conservative   54;  Mismatches   98;  Indels   78;  Gaps   18;

Qy         89 LIDFEGKKDVAQIFNN--------ILRRQI--GTRTPTVEYICT--QQNILFMLLK---G 133
Db        401 LLNYIFKSDQTYFYHNNFIDEYKQKICKQIKCSTKKNDISHIITSRKENHLFHVQKLENN 460

Qy        134 YESPEIALNCGIMLRECIRHEPLAKIILWSEQFYDFFRYVEMSTFDIASDAFATFKDLLT 193
Db        461 YKHPNI--------NKQLRKTKILKYVY--NYFKEFINNVINTKFGKIYRKFFPRKHILN 510

Qy        194 R-HKL----------------LSAEFLEQH-YDRFFSEYE---------KLLHSENYV 224
Db        511 KIHKIFKIIRLQIIKKYRIINIRMNRKFIKQKVYDTFFKNYDFLSFSFKTYKIINFMVYI 570

Qy        225 TKR-QSLKLLGELLLDRHNFTIMTKYISK------PENLKLMMNLLRDKSRNIQFEAFHV 277
Db        571 TKKCIPIKLLG----SKHNFKIFLKNVKKFLLFNYKESFSLNQVMKNIKVKNI-FQK-KI 624

Qy        278 FKVFVANPNKTQPILD------ILLKNQAKLIEFLS---KFQNDRTDCMSSS--VPTT 324
Db        625 SKYNIKNRILLKNIFDNNYENKILHRNNKEIITNINDNIKIYNKKNDNLNNSFKIKTT 682


RESULT 15
US-09-134-001C-4053
; Sequence 4053, Application US/09134001C
; Patent No. 6380370
; GENERAL INFORMATION:
;  APPLICANT: Lynn Doucette-Stamm et al
;  TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
;  TITLE OF INVENTION:  EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
;  FILE REFERENCE: GTC-007
;  CURRENT APPLICATION NUMBER: US/09/134,001C
;  CURRENT FILING DATE:  1998-08-13
;  PRIOR APPLICATION NUMBER: US 60/064,964
;  PRIOR FILING DATE: 1997-11-08
;  PRIOR APPLICATION NUMBER: US 60/055,779
;  PRIOR FILING DATE: 1997-08-14
;  NUMBER OF SEQ ID NOS: 5674
; SEQ ID NO 4053
;   LENGTH: 533
;   TYPE: PRT
;   ORGANISM: Staphylococcus epidermidis
US-09-134-001C-4053

  Query Match             6.2%;  Score 110.5;  DB 3;  Length 533;
  Best Local Similarity   22.5%;  Pred. No. 0.029;
  Matches   73;  Conservative   57;  Mismatches  108;  Indels   87;  Gaps   18;

Qy         14 DIVKNLKESMAVL----EKQDISDKKAEKA-TEEVSKNLVAMKEILYGTNEK-------- 60
Db        237 DILRQFSDSIDVLITGHQHRDIATIKNQTAIIQPGSKGTKVGKIVIEYTHDKKVLIKECN 296

Qy         61 ----------EPQTEAVAQLAQELYNSGLLSTLVADLQLI-----DFEGKKD---VAQIF 102
Db        297 LMNVHNSTSFKPNDEDIA-LRNQL--EDWLDTQIAELPYAMRINNSFEARKSPHAFVNLL 353

Qy        103 NNILRRQIGTRTPTVEYICTQQNILFMLLKGYESPEIALNCGIMLRECIRHEPLAKIILW 162
Db        354 NYILLEKSG-----ADIACT---ALFDSANGFDEK-------VTMRDIINNYP------- 391

Qy        163 SEQFYDFFRYVEMSTFDI--ASDAFATFKDLLTRHKLLSAEFLE---QHYD-RFFSEYEK 216
Db        392 ---FPNTFKVIELSGKDIKLAIERSASYFDIVNHKITVNKEFLEPKPQHFNYDIFAGIQY 448

Qy        217 LLHSENYVTKRQSLKLLGELLLDRHN-FTIMTK------------YISKP-------ENL 256
Db        449 TIHVSHPYGERVSDLLINDAPLQSDQIYTICVNNYRAVGGGNYDMYVNKPVIKDIQIEGA 508

Qy        257 KLMMNLL--RDKSRNIQFEAFHVFK 279
Db        509 QLLIDYLSHNDLSQIPQVIDFNVVK 533


Search completed: October 22, 2004, 09:13:52
Job time : 42 secs
```

GenCore version 5.1.6

OM protein - protein search, using sw model

Run on: October 22, 2004, 08:55:18 ; Search time 157 Seconds
(without alignments)
799.715 Million cell updates/sec

Title: US-10-089-688-2
Perfect score: 1781
Sequence: 1 MPPPFGKSHKSPADIVKNLK.....RVKPRTRGIRDLKRPAQQEA 350

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 2002273 seqs, 358729299 residues

Total number of hits satisfying chosen parameters: 2002273

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :
A_Geneseq_23Sep04:*
1: geneseqp1980s:*
2: geneseqp1990s:*
3: geneseqp2000s:*
4: geneseqp2001s:*
5: geneseqp2002s:*
6: geneseqp2003as:*
7: geneseqp2003bs:*
8: geneseqp2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	1781	100.0	350	4	AAB20387	Aab20387 Human acu
2	1635.5	91.8	341	4	AAB48970	Aab48970 Human ANI
3	1635.5	91.8	341	7	ADJ69105	Adj69105 Human hea
4	1635.5	91.8	496	4	AAE10858	Aae10858 Gal4-huma
5	1635.5	91.8	552	4	AAE10859	Aae10859 LexA-huma
6	1614.5	90.7	341	3	AAY94248	Aay94248 Mouse cal
7	1597.5	89.7	354	4	ABG23844	Abg23844 Novel hum
8	1297.5	72.9	337	3	AAY94247	Aay94247 Human cal
9	1297.5	72.9	337	4	AAB82090	Aab82090 Human Acu
10	1297.5	72.9	337	4	AAM39078	Aam39078 Human pol
11	1153.5	64.8	289	4	AAB94139	Aab94139 Human pro
12	1123	63.1	339	4	ABB60392	Abb60392 Drosophil
13	1117	62.7	339	3	AAY94249	Aay94249 Drosophil
14	1054.5	59.2	377	3	AAY94250	Aay94250 C. elegan
15	888.5	49.9	237	4	AAM40864	Aam40864 Human pol
16	705.5	39.6	343	3	AAG45273	Aag45273 Arabidops
17	674	37.8	326	3	AAG51053	Aag51053 Arabidops
18	674	37.8	400	3	AAG51052	Aag51052 Arabidops
19	674	37.8	504	3	AAG51051	Aag51051 Arabidops
20	673.5	37.8	300	3	AAG23886	Aag23886 Arabidops
21	669.5	37.6	345	3	AAG05089	Aag05089 Arabidops
22	666.5	37.4	300	3	AAG45274	Aag45274 Arabidops
23	666.5	37.4	300	3	AAG30714	Aag30714 Arabidops
24	666.5	37.4	305	3	AAG30713	Aag30713 Arabidops
25	658.5	37.0	290	3	AAG23887	Aag23887 Arabidops
26	635	35.7	320	3	AAG05090	Aag05090 Arabidops
27	523.5	29.4	213	3	AAG23888	Aag23888 Arabidops
28	518.5	29.1	213	3	AAG30715	Aag30715 Arabidops
29	518.5	29.1	213	3	AAG45275	Aag45275 Arabidops
30	465.5	26.1	154	3	AAG41151	Aag41151 Zea mays
31	460	25.8	197	3	AAG05091	Aag05091 Arabidops
32	448.5	25.2	148	3	AAG41152	Aag41152 Zea mays
33	430.5	24.2	139	3	AAG41153	Aag41153 Zea mays
34	276	15.5	639	4	ABG25372	Abg25372 Novel hum
35	271	15.2	383	4	ABG23843	Abg23843 Novel hum
36	269	15.1	236	5	ABP02921	Abp02921 Human ORF
37	209	11.7	135	5	ABP34081	Abp34081 Human ORF
38	127.5	7.2	121	3	AAG33914	Aag33914 Arabidops
39	126	7.1	1182	6	ABU25521	Abu25521 Protein e
40	116.5	6.5	888	5	AAU83013	Aau83013 Human hom
41	116.5	6.5	972	6	ABJ19394	Abj19394 Human int
42	113	6.3	541	7	ADH87535	Adh87535 Enterococ
43	113	6.3	2184	4	AAE00425	Aae00425 P. falcip
44	111.5	6.3	790	7	ADJ70372	Adj70372 Human hea
45	110.5	6.2	533	5	ABP39208	Abp39208 Staphyloc

ALIGNMENTS

RESULT 1
AAB20387
ID AAB20387 standard; protein; 350 AA.
XX
AC AAB20387;
XX
DT 11-JUN-2001 (first entry)
XX
DE Human acute neuronal induced calcium binding protein ANIC-BP-1B.
XX
KW Acute neuronal induced calcium binding protein; ANIC-BP-1B;
KW spice variant; human; stroke; head trauma; Parkinson's disease;
KW Alzheimer's disease; multiple sclerosis; spinal cord injury;
KW cerebroprotective; antiparkinsonian; nootropic; neuroprotective; therapy;
KW diagnosis; vaccine.
XX
OS Homo sapiens.
XX
PN WO200125423-A1.
XX
PD 12-APR-2001.
XX
PF 28-SEP-2000; 2000WO-EP009475.
XX
PR 04-OCT-1999; 99EP-00119113.
XX
PA (MERE) MERCK PATENT GMBH.
XX
PI Duecker K, Den Daas I;
XX
DR WPI; 2001-266306/27.
DR N-PSDB; AAF30688.
XX
PT Novel human acute neuronal induced calcium-binding protein like protein
PT splice variant, useful for treating stroke, acute head trauma,
PT Parkinson's disease, Alzheimer's disease multiple sclerosis, spinal cord
PT injury.
XX
PS Claim 2; Page 44-45; 49pp; English.
XX
CC The present sequence is that of a novel human acute neuronal induced
CC calcium binding protein-like protein splice variant, ANIC-NP-1B. The
CC protein shows homology to other members of the calcium binding protein
CC family, including ANIC-BP, a protein discovered by mRNA differential
CC display that is upregulated in a rat model of head trauma. ANIC-BP and
CC ANIC-BP-1B differ in their C-terminal portions. The variant protein could
CC serve as a novel drug target. The invention provides ANIC-BP-1B
CC polynucleotides (see AAF30688) and polypeptides, expression vectors, host

```
CC   cells and antibodies, as well as methods for producing the protein and
CC   for treating or preventing disorders associated with expression of the
CC   protein by inhibiting or activating the action of ANIC-BP-1B. Diseases
CC   that may be treated include stroke and acute head trauma, Parkinson's
CC   disease, Alzheimer's disease, multiple sclerosis and spinal cord injury.
CC   The polynucleotides and polypeptides can also be used in diagnostic
CC   assays and in vaccines, and to identify agonists and antagonists useful
CC   for treating conditions associated with ANIC-BP-1B imbalance
XX
SQ   Sequence 350 AA;

  Query Match             100.0%;  Score 1781;  DB 4;  Length 350;
  Best Local Similarity   100.0%;  Pred. No. 6.8e-157;
  Matches  350;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 MPFPFGKSHKSPADIVKNLKESMAVLEKQDISDKKAEKATEEVSKNLVAMKEILYGTNEK 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MPFPFGKSHKSPADIVKNLKESMAVLEKQDISDKKAEKATEEVSKNLVAMKEILYGTNEK 60

Qy         61 EPQTEAVAQLAQELYNSGLLSTLVADLQLIDFEGKKDVAQIFNNILRRQIGTRTPTVEYI 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 EPQTEAVAQLAQELYNSGLLSTLVADLQLIDFEGKKDVAQIFNNILRRQIGTRTPTVEYI 120

Qy        121 CTQQNILFMLLKGYESPEIALNCGIMLRECIRHEPLAKIILWSEQFYDFFRYVEMSTFDI 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CTQQNILFMLLKGYESPEIALNCGIMLRECIRHEPLAKIILWSEQFYDFFRYVEMSTFDI 180

Qy        181 ASDAFATFKDLLTRHKLLSAEFLEQHYDRFFSEYEKLLHSENYVTKRQSLKLLGELLLDR 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 ASDAFATFKDLLTRHKLLSAEFLEQHYDRFFSEYEKLLHSENYVTKRQSLKLLGELLLDR 240

Qy        241 HNFTIMTKYISKPENLKLMMNLLRDKSRNIQFEAFHVFKVFVANPNKTQPILDILLKNQA 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 HNFTIMTKYISKPENLKLMMNLLRDKSRNIQFEAFHVFKVFVANPNKTQPILDILLKNQA 300

Qy        301 KLIEFLSKFQNDRTDCMSSSVPTTNSRVDLRVKPRTRGIRDLKRPAQQEA 350
              ||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 KLIEFLSKFQNDRTDCMSSSVPTTNSRVDLRVKPRTRGIRDLKRPAQQEA 350


RESULT 2
AAB48970
ID   AAB48970 standard; protein; 341 AA.
XX
AC   AAB48970;
XX
DT   27-MAR-2001  (first entry)
XX
DE   Human ANIC-BP (acute neuronal induced calcium-binding protein).
XX
KW   Human; acute neuronal induced calcium-binding protein; ANIC-BP;
KW   Mo25 homologue; HymA homologue; drug screening; stroke;
KW   acute head trauma; multiple sclerosis; spinal cord injury; vaccine;
KW   cerebroprotective; neuroprotective.
XX
OS   Homo sapiens.
XX
PN   WO200078947-A1.
XX
PD   28-DEC-2000.
XX
PF   14-JUN-2000; 2000WO-EP005457.
XX
PR   22-JUN-1999;   99EP-00112024.
XX
PA   (MERE ) MERCK PATENT GMBH.
XX
PI   Den Daas I,  Fischer V,  Seyfried C,  Von Melchner L;
XX
DR   WPI; 2001-102721/11.
DR   N-PSDB; AAC91772.
XX
PT   Novel acute neuronal induced calcium binding protein, useful for treating
PT   acute head trauma, stroke, multiple sclerosis and spinal cord injury.
XX
PS   Claim 2; Page 37; 50pp; English.
XX
CC   The invention relates to human acute neuronal induced calcium-binding
CC   protein (ANIC-BP) and to nucleic acid encoding it. The invention also
CC   relates to expression systems and recombinant host cells comprising ANIC-
CC   BP DNA, the recombinant production of ANIC-BP, antibodies specific for
CC   ANIC-BP, fusion proteins comprising ANIC-BP and an immunoglobulin Fc
CC   region, and methods of screening for modulators of ANIC-BP function. ANIC
CC   -BP has homology and structural similarity to HymA and Mo25 proteins.
CC   ANIC-BP proteins and nucleotides are useful for treating stroke and acute
CC   head trauma, multiple sclerosis and spinal cord injury. ANIC-BP proteins
CC   are useful in screening assays, for identifying membrane bound or soluble
CC   receptors, and also in vaccines. ANIC-BP nucleotides are useful as
CC   diagnostic reagents, as tools for tissue expression studies, for
CC   chromosome localisation studies, as genetic vaccines, and in the
CC   generation of transgenic animals. The present sequence represents human
CC   ANIC-BP
XX
SQ   Sequence 341 AA;

  Query Match             91.8%;  Score 1635.5;  DB 4;  Length 341;
  Best Local Similarity   94.0%;  Pred. No. 2.3e-143;
  Matches  329;  Conservative    1;  Mismatches   11;  Indels    9;  Gaps    2;

Qy          1 MPFPFGKSHKSPADIVKNLKESMAVLEKQDISDKKAEKATEEVSKNLVAMKEILYGTNEK 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MPFPFGKSHKSPADIVKNLKESMAVLEKQDISDKKAEKATEEVSKNLVAMKEILYGTNEK 60

Qy         61 EPQTEAVAQLAQELYNSGLLSTLVADLQLIDFEGKKDVAQIFNNILRRQIGTRTPTVEYI 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 EPQTEAVAQLAQELYNSGLLSTLVADLQLIDFEGKKDVAQIFNNILRRQIGTRTPTVEYI 120

Qy        121 CTQQNILFMLLKGYESPEIALNCGIMLRECIRHEPLAKIILWSEQFYDFFRYVEMSTFDI 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CTQQNILFMLLKGYESPEIALNCGIMLRECIRHEPLAKIILWSEQFYDFFRYVEMSTFDI 180

Qy        181 ASDAFATFKDLLTRHKLLSAEFLEQHYDRFFSEYEKLLHSENYVTKRQSLKLLGELLLDR 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 ASDAFATFKDLLTRHKLLSAEFLEQHYDRFFSEYEKLLHSENYVTKRQSLKLLGELLLDR 240

Qy        241 HNFTIMTKYISKPENLKLMMNLLRDKSRNIQFEAFHVFKVFVANPNKTQPILDILLKNQA 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 HNFTIMTKYISKPENLKLMMNLLRDKSRNIQFEAFHVFKVFVANPNKTQPILDILLKNQA 300

Qy        301 KLIEFLSKFQNDRTDCMSSSVPTTNSRVDLRVKPRTRGIRDLKRPAQQEA 350
              |||||||||||||:     |      ||     ||||||||||||
Db        301 KLIEFLSKFQNDRTE-----DEQFNDEKTYLVKQ----IRDLKRPAQQEA 341


RESULT 3
ADJ69105
ID   ADJ69105 standard; protein; 341 AA.
XX
AC   ADJ69105;
XX
DT   06-MAY-2004  (first entry)
XX
DE   Human heat mitochondrial protein as a therapeutic target SeqID911.
XX
KW   mitochondrial; human; screening assay; diabetes mellitus;
KW   Huntington's disease; osteoarthritis;
KW   Leber's hereditary optic neuropathy; LHON;
KW   mitochondrial encephalopathy lactic acidosis and stroke; MELAS;
KW   myoclonic epilepsy ragged red fibre syndrome; MERRF; cancer;
KW   neuroprotective; nootropic; antidiabetic; anticonvulsant; antiarthritic;
KW   osteopathic; ophthalmological; cytostatic.
XX
OS   Homo sapiens.
XX
PN   WO2003087768-A2.
XX
```

PD 23-OCT-2003.
XX
PF 04-APR-2003; 2003WO-US010870.
XX
PR 12-APR-2002; 2002US-0372843P.
PR 17-JUN-2002; 2002US-0389987P.
PR 20-SEP-2002; 2002US-0412418P.
XX
PA (MITO-) MITOKOR.
PA (BUCK-) BUCK INST AGE RES.
XX
PI Ghosh SS, Fahy ED, Zhang B, Gibson BW, Taylor SW, Glenn GM;
PI Warnock DE;
XX
DR WPI; 2003-845369/78.
XX
PT Identifying a mitochondrial target for drug screening assays and for
PT treating diseases associated with altered mitochondrial function,
PT comprises detecting a modified polypeptide in a sample and correlating
PT with the disease.
XX
PS Claim 1; SEQ ID NO 911; 180pp; English.
XX
CC This invention relates to novel mitochondrial targets that can be used
CC for therapeutic intervention in treating a disease associated with
CC altered mitochondrial function. Specifically, it refers to a method for
CC identifying proteins of the human heart mitochondrial proteome that are
CC useful for drug screening assays, as well as therapeutic targets. The
CC present invention describes a method for identifying such proteins that
CC can be used in the treatment of various diseases associated with altered
CC mitochondrial function including diabetes mellitus, Huntington's disease,
CC osteoarthritis, Leber's hereditary optic neuropathy (LHON), mitochondrial
CC encephalopathy lactic acidosis and stroke (MELAS), myoclonic epilepsy
CC ragged red fibre syndrome (MERRF) or cancer. Accordingly, these
CC compositions have neuroprotective, nootropic, antidiabetic,
CC anticonvulsant, antiarthritic, osteopathic, ophthalmological and
CC cytostatic activities. This polypeptide sequence is a human heart
CC mitochondrial protein of the invention.
XX
SQ Sequence 341 AA;

Query Match 91.8%; Score 1635.5; DB 7; Length 341;
Best Local Similarity 94.0%; Pred. No. 2.3e-143;
Matches 329; Conservative 1; Mismatches 11; Indels 9; Gaps 2;

```
Qy          1 MPFPPFGKSHKSPADIVKNLKESMAVLEKQDISDKKAEKATEEVSKNLVAMKEILYGTNEK 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MPFPPFGKSHKSPADIVKNLKESMAVLEKQDISDKKAEKATEEVSKNLVAMKEILYGTNEK 60

Qy         61 EPQTEAVAQLAQELYNSGLLSTLVADLQLIDFEGKKDVAQIFNNILRRQIGTRTPTVEYI 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 EPQTEAVAQLAQELYNSGLLSTLVADLQLIDFEGKKDVAQIFNNILRRQIGTRTPTVEYI 120

Qy        121 CTQQNILFMLLKGYESPEIALNCGIMLRECIRHEPLAKIILWSEQFYDFFRYVEMSTFDI 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CTQQNILFMLLKGYESPEIALNCGIMLRECIRHEPLAKIILWSEQFYDFFRYVEMSTFDI 180

Qy        181 ASDAFATFKDLLTRHKLLSAEFLEQHYDRFFSEYEKLLHSENYVTKRQSLKLLGELLLDR 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 ASDAFATFKDLLTRHKLLSAEFLEQHYDRFFSEYEKLLHSENYVTKRQSLKLLGELLLDR 240

Qy        241 HNFTIMTKYISKPENLKLMMNLLRDKSRNIQFEAFHVFKVFVANPNKTQPILDILLKNQA 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 HNFTIMTKYISKPENLKLMMNLLRDKSRNIQFEAFHVFKVFVANPNKTQPILDILLKNQA 300

Qy        301 KLIEFLSKFQNDRTDCMSSSVPTTNSRVDLRVKPRTRGIRDLKRPAQQEA 350
              |||||||||||||:          |       ||     ||||||||||
Db        301 KLIEFLSKFQNDRTE-----DEQFNDEKTYLVKQ----IRDLKRPAQQEA 341
```

RESULT 4
AAE10858
ID AAE10858 standard; protein; 496 AA.
XX
AC AAE10858;
XX
DT 18-DEC-2001 (first entry)
XX
DE Gal4-human ANIC-BP-1 fusion protein.
XX
KW Human; acute neuronal induced calcium binding protein type 1 ligand;
KW ANIC-BP-1; human disease; stroke; head trauma; multiple sclerosis;
KW Parkinson's disease; Alzheimer's disease; spinal cord injury; vaccine;
KW gene therapy; fusion protein; Gal4 protein.
XX
OS Homo sapiens.
OS Unidentified.
OS Chimeric.
XX
PN WO200170771-A2.
XX
PD 27-SEP-2001.
XX
PF 20-MAR-2001; 2001WO-EP003149.
XX
PR 21-MAR-2000; 2000EP-00106110.
XX
PA (MERE) MERCK PATENT GMBH.
XX
PI Den Daas I, Duecker K, Hock B;
XX
DR WPI; 2001-607519/69.
XX
PT Novel acute neuronal induced calcium binding protein type 1 ligand
PT polypeptides, useful in the treatment of stroke, head trauma, multiple
PT sclerosis, Parkinson's disease, Alzheimer's disease and spinal cord
PT injury.
XX
PS Disclosure; Page 42-44; 46pp; English.
XX
CC The invention relates to human acute neuronal induced calcium binding
CC protein type 1 (ANIC-BP-1) ligand polypeptides and polynucleotides.
CC Sequences of the invention are useful for treating human diseases
CC including stroke, head trauma, multiple sclerosis, Parkinson's disease,
CC Alzheimer's disease and spinal cord injury. They are also useful as
CC vaccines. ANIC-BP-1 ligands are useful for identifying membrane bound
CC soluble receptors. Polynucleotides of the invention are useful as
CC diagnostic reagents, for chromosome localization studies, and as valuable
CC tools for tissue expression studies. They are also useful in gene
CC therapy. The present sequence is Gal4-human ANIC-BP-1 fusion protein
CC comprising the Gal4 protein and a C-terminally linked human ANIC-BP-1
CC protein
XX
SQ Sequence 496 AA;

Query Match 91.8%; Score 1635.5; DB 4; Length 496;
Best Local Similarity 94.0%; Pred. No. 3.9e-143;
Matches 329; Conservative 1; Mismatches 11; Indels 9; Gaps 2;

```
Qy          1 MPFPPFGKSHKSPADIVKNLKESMAVLEKQDISDKKAEKATEEVSKNLVAMKEILYGTNEK 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        156 MPFPPFGKSHKSPADIVKNLKESMAVLEKQDISDKKAEKATEEVSKNLVAMKEILYGTNEK 215

Qy         61 EPQTEAVAQLAQELYNSGLLSTLVADLQLIDFEGKKDVAQIFNNILRRQIGTRTPTVEYI 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        216 EPQTEAVAQLAQELYNSGLLSTLVADLQLIDFEGKKDVAQIFNNILRRQIGTRTPTVEYI 275

Qy        121 CTQQNILFMLLKGYESPEIALNCGIMLRECIRHEPLAKIILWSEQFYDFFRYVEMSTFDI 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        276 CTQQNILFMLLKGYESPEIALNCGIMLRECIRHEPLAKIILWSEQFYDFFRYVEMSTFDI 335

Qy        181 ASDAFATFKDLLTRHKLLSAEFLEQHYDRFFSEYEKLLHSENYVTKRQSLKLLGELLLDR 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        336 ASDAFATFKDLLTRHKLLSAEFLEQHYDRFFSEYEKLLHSENYVTKRQSLKLLGELLLDR 395

Qy        241 HNFTIMTKYISKPENLKLMMNLLRDKSRNIQFEAFHVFKVFVANPNKTQPILDILLKNQA 300
```

```
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        396 HNFTIMTKYISKPENLKLMMNLLRDKSRNIQFEAFHVFKVFVANPNKTQPILDILLKNQA 455

Qy        301 KLIEFLSKFQNDRTDCMSSSVPTTNSRVDLRVKPRTRGIRDLKRPAQQEA 350
                ||||||||||||||:         |       ||     ||||||||||||
Db        456 KLIEFLSKFQNDRTE-----DEQFNDEKTYLVKQ----IRDLKRPAQQEA 496


RESULT 5
AAE10859
ID   AAE10859 standard; protein; 552 AA.
XX
AC   AAE10859;
XX
DT   18-DEC-2001  (first entry)
XX
DE   LexA-human ANIC-BP-1 fusion protein.
XX
KW   Human; acute neuronal induced calcium binding protein type 1 ligand;
KW   ANIC-BP-1; human disease; stroke; head trauma; multiple sclerosis;
KW   Parkinson's disease; Alzheimer's disease; spinal cord injury; vaccine;
KW   gene therapy; fusion protein; LexA protein.
XX
OS   Homo sapiens.
OS   Unidentified.
OS   Chimeric.
XX
FH   Key             Location/Qualifiers
FT   Region          1. .202
FT                   /note= "LexA protein"
FT   Region          203. .552
FT                   /note= "Human ANIC-BP-1 protein"
XX
PN   WO200170771-A2.
XX
PD   27-SEP-2001.
XX
PF   20-MAR-2001; 2001WO-EP003149.
XX
PR   21-MAR-2000; 2000EP-00106110.
XX
PA   (MERE ) MERCK PATENT GMBH.
XX
PI   Den Daas I,  Duecker K,  Hock B;
XX
DR   WPI; 2001-607519/69.
XX
PT   Novel acute neuronal induced calcium binding protein type 1 ligand
PT   polypeptides, useful in the treatment of stroke, head trauma, multiple
PT   sclerosis, Parkinson's disease, Alzheimer's disease and spinal cord
PT   injury.
XX
PS   Disclosure; Page 44-46; 46pp; English.
XX
CC   The invention relates to human acute neuronal induced calcium binding
CC   protein type 1 (ANIC-BP-1) ligand polypeptides and polynucleotides.
CC   Sequences of the invention are useful for treating human diseases
CC   including stroke, head trauma, multiple sclerosis, Parkinson's disease,
CC   Alzheimer's disease and spinal cord injury. They are also useful as
CC   vaccines. ANIC-BP-1 ligands are useful for identifying membrane bound
CC   soluble receptors. Polynucleotides of the invention are useful as
CC   diagnostic reagents, for chromosome localization studies, and as valuable
CC   tools for tissue expression studies. They are also useful in gene
CC   therapy. The present sequence is LexA-human ANIC-BP-1 fusion protein
CC   comprising the LexA protein and a C-terminally linked human ANIC-BP-1
CC   protein
XX
SQ   Sequence 552 AA;

  Query Match             91.8%;  Score 1635.5;  DB 4;  Length 552;
  Best Local Similarity   94.0%;  Pred. No. 4.5e-143;
  Matches  329;  Conservative    1;  Mismatches   11;  Indels    9;  Gaps    2;

Qy          1 MPFPPFGKSHKSPADIVKNLKESMAVLEKQDISDKKAEKATEEVSKNLVAMKEILYGTNEK 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        212 MPFPPFGKSHKSPADIVKNLKESMAVLEKQDISDKKAEKATEEVSKNLVAMKEILYGTNEK 271

Qy         61 EPQTEAVAQLAQELYNSGLLSTLVADLQLIDFEGKKDVAQIFNNILRRQIGTRTPTVEYI 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        272 EPQTEAVAQLAQELYNSGLLSTLVADLQLIDFEGKKDVAQIFNNILRRQIGTRTPTVEYI 331

Qy        121 CTQQNILFMLLKGYESPEIALNCGIMLRECIRHEPLAKIILWSEQFYDFFRYVEMSTFDI 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        332 CTQQNILFMLLKGYESPEIALNCGIMLRECIRHEPLAKIILWSEQFYDFFRYVEMSTFDI 391

Qy        181 ASDAFATFKDLLTRHKLLSAEFLEQHYDRFFSEYEKLLHSENYVTKRQSLKLLGELLLDR 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        392 ASDAFATFKDLLTRHKLLSAEFLEQHYDRFFSEYEKLLHSENYVTKRQSLKLLGELLLDR 451

Qy        241 HNFTIMTKYISKPENLKLMMNLLRDKSRNIQFEAFHVFKVFVANPNKTQPILDILLKNQA 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        452 HNFTIMTKYISKPENLKLMMNLLRDKSRNIQFEAFHVFKVFVANPNKTQPILDILLKNQA 511

Qy        301 KLIEFLSKFQNDRTDCMSSSVPTTNSRVDLRVKPRTRGIRDLKRPAQQEA 350
              ||||||||||||||:         |       ||     ||||||||||||
Db        512 KLIEFLSKFQNDRTE-----DEQFNDEKTYLVKQ----IRDLKRPAQQEA 552


RESULT 6
AAY94248
ID   AAY94248 standard; protein; 341 AA.
XX
AC   AAY94248;
XX
DT   10-AUG-2000  (first entry)
XX
DE   Mouse calcium binding protein MO25.
XX
KW   Mouse; calcium binding protein; cancer; inflammation; MO25; CBP;
KW   reproductive disorder; autoimmune disorder; developmental disorder;
KW   seizure disorder; immune disorder; infection.
XX
OS   Mus sp.
XX
PN   WO200029580-A1.
XX
PD   25-MAY-2000.
XX
PF   12-NOV-1999;   99WO-US027027.
XX
PR   13-NOV-1998;   98US-00190965.
XX
PA   (INCY-) INCYTE PHARM INC.
XX
PI   Tang YT,  Guegler KJ,  Corley NC,  Gorgone GA;
XX
DR   WPI; 2000-387793/33.
XX
PT   Human hCBP protein, and the nucleic acid encoding it, useful for e.g.
PT   diagnosis, prevention and treatment of cancers, immune, developmental or
PT   reproductive disorders.
XX
PS   Disclosure; Page 66-67; 72pp; English.
XX
CC   The present sequence is the mouse calcium binding protein MO25. It was
CC   used in a sequence alignment to identify human calcium binding protein
CC   hCBP. The hCBP protein and the gene encoding it are useful for the
CC   diagnosis and treatment of the following types of disorder: cancers (such
CC   as adenocarcinomas), reproductive disorders (such as infertility,
CC   ovulatory defects, endometriosis, disruptions of the oestrus and
CC   menstrual cycles, polycystic ovary syndrome and ovarian
CC   hyperstimulation), autoimmune disorders (such as benign prostatic
CC   hyperplasia and prostatitis), developmental disorders (such as Cushing's
CC   syndrome, muscular dystrophy and gonadal dysgenesis), hereditary
CC   neuropathies, seizure disorders, immune disorders (such as AIDS,
CC   allergies, anaemia, asthma, atherosclerosis, cholecystitis, Crohn's
```

```
CC   disease, diabetes, Graves' disease, multiple sclerosis, psoriasis,
CC   rheumatoid arthritis, scleroderma, Sjogren's syndrome and ulcerative
CC   colitis), and viral, bacterial, fungal, parasitic, protozoal and
CC   helminthic infections
XX
SQ   Sequence 341 AA;

  Query Match             90.7%;  Score 1614.5;  DB 3;  Length 341;
  Best Local Similarity   92.9%;  Pred. No. 2e-141;
  Matches  325;  Conservative    2;  Mismatches   14;  Indels    9;  Gaps    2;

Qy          1 MPFPPFGKSHKSPADIVKNLKESMAVLEKQDISDKKAEKATEEVSKNLVAMKEILYGTNEK 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MPFPPFGKSHKSPADIVKNLKESMAVLEKQDISDKKAEKATEEVSKNLVAMKEILYGTNEK 60

Qy         61 EPQTEAVAQLAQELYNSGLLSTLVADLQLIDFEGKKDVAQIFNNILRRQIGTRTPTVEYI 120
              |||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||||
Db         61 EPQTEAVAQLAQELYNSGLLGTLVADLQLIDFEGKKDVAQIFNNILRRQIGTRTPTVEYI 120

Qy        121 CTQQNILFMLLKGYESPEIALNCGIMLRECIRHEPLAKIILWSEQFYDFFRYVEMSTFDI 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CTQQNILFMLLKGYESPEIALNCGIMLRECIRHEPLAKIILWSEQFYDFFRYVEMSTFDI 180

Qy        181 ASDAFATFKDLLTRHKLLSAEFLEQHYDRFFSEYEKLLHSENYVTKRQSLKLLGELLLDR 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 ASDAFATFKDLLTRHKLLSAEFLEQHYDRFFSEYEKLLHSENYVTKRQSLKLLGELLLDR 240

Qy        241 HNFTIMTKYISKPENLKLMMNLLRDKSRNIQFEAFHVFKVFVANPNKTQPILDILLKNQA 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 HNFTIMTKYISKPENLKLMMNLLRDKSRNIQFEAFHVFKVFVANPNKTQPILDILLKNQT 300

Qy        301 KLIEFLSKFQNDRTDCMSSSVPTTNSRVDLRVKPRTRGIRDLKRPAQQEA 350
              |||||||||||||:        |       ||      ||:||| |||||
Db        301 KLIEFLSKFQNDRTE-----DEQFNDEKTYLVKQ----IRNLKRAAQQEA 341


RESULT 7
ABG23844
ID   ABG23844 standard; protein; 354 AA.
XX
AC   ABG23844;
XX
DT   18-FEB-2002  (first entry)
XX
DE   Novel human diagnostic protein #23835.
XX
KW   Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW   food supplement; medical imaging; diagnostic; genetic disorder.
XX
OS   Homo sapiens.
XX
PN   WO200175067-A2.
XX
PD   11-OCT-2001.
XX
PF   30-MAR-2001; 2001WO-US008631.
XX
PR   31-MAR-2000; 2000US-00540217.
PR   23-AUG-2000; 2000US-00649167.
XX
PA   (HYSE-) HYSEQ INC.
XX
PI   Drmanac RT,  Liu C,  Tang YT;
XX
DR   WPI; 2001-639362/73.
DR   N-PSDB; AAS88031.
XX
PT   New isolated polynucleotide and encoded polypeptides, useful in
PT   diagnostics, forensics, gene mapping, identification of mutations
PT   responsible for genetic disorders or other traits and to assess
PT   biodiversity.
XX
PS   Claim 20; SEQ ID NO 54203; 103pp; English.
XX
CC   The invention relates to isolated polynucleotide (I) and polypeptide (II)
CC   sequences. (I) is useful as hybridisation probes, polymerase chain
CC   reaction (PCR) primers, oligomers, and for chromosome and gene mapping,
CC   and in recombinant production of (II). The polynucleotides are also used
CC   in diagnostics as expressed sequence tags for identifying expressed
CC   genes. (I) is useful in gene therapy techniques to restore normal
CC   activity of (II) or to treat disease states involving (II). (II) is
CC   useful for generating antibodies against it, detecting or quantitating a
CC   polypeptide in tissue, as molecular weight markers and as a food
CC   supplement. (II) and its binding partners are useful in medical imaging
CC   of sites expressing (II). (I) and (II) are useful for treating disorders
CC   involving aberrant protein expression or biological activity. The
CC   polypeptide and polynucleotide sequences have applications in
CC   diagnostics, forensics, gene mapping, identification of mutations
CC   responsible for genetic disorders or other traits to assess biodiversity
CC   and to produce other types of data and products dependent on DNA and
CC   amino acid sequences. ABG00010-ABG30377 represent novel human diagnostic
CC   amino acid sequences of the invention. Note: The sequence data for this
CC   patent did not appear in the printed specification, but was obtained in
CC   electronic format directly from WIPO at
CC   ftp.wipo.int/pub/published_pct_sequences
XX
SQ   Sequence 354 AA;

  Query Match             89.7%;  Score 1597.5;  DB 4;  Length 354;
  Best Local Similarity   92.0%;  Pred. No. 8.2e-140;
  Matches  322;  Conservative    2;  Mismatches   17;  Indels    9;  Gaps    2;

Qy          1 MPFPPFGKSHKSPADIVKNLKESMAVLEKQDISDKKAEKATEEVSKNLVAMKEILYGTNEK 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         14 MPFPPFGKSHKSPADIVKNLKESMAVLEKQDISDKKAEKATEEVSKNLVAMKEILYGTNEK 73

Qy         61 EPQTEAVAQLAQELYNSGLLSTLVADLQLIDFEGKKDVAQIFNNILRRQIGTRTPTVEYI 120
              :||||| |||||||||||||| ||||||||||||||||||||||||||||||||||||||||
Db         74 DPQTEAGAQLAQELYNSGLLITLVADLQLIDFEGKKDVAQIFNNILRRQIGTRTPTVEYI 133

Qy        121 CTQQNILFMLLKGYESPEIALNCGIMLRECIRHEPLAKIILWSEQFYDFFRYVEMSTFDI 180
              |||||||||||||||||||||||||||||||||| ||||||||||||||||||||||||||
Db        134 CTQQNILFMLLKGYESPEIALNCGIMLRECIRHEPLGKIILWSEQFYDFFRYVEMSTFDI 193

Qy        181 ASDAFATFKDLLTRHKLLSAEFLEQHYDRFFSEYEKLLHSENYVTKRQSLKLLGELLLDR 240
              |||||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        194 ASDAFATFKGLLTRHKLLSAEFLEQHYDRFFSEYEKLLHSENYVTKRQSLKLLGELLLDR 253

Qy        241 HNFTIMTKYISKPENLKLMMNLLRDKSRNIQFEAFHVFKVFVANPNKTQPILDILLKNQA 300
              |||||||||||||| |||||||||||||||||||||||||||| ||||||||||||||||||
Db        254 HNFTIMTKYISKPVNLKLMMNLLRDKSRNIQFEAFHVFKAFVANPNKTQPILDILLKNQA 313

Qy        301 KLIEFLSKFQNDRTDCMSSSVPTTNSRVDLRVKPRTRGIRDLKRPAQQEA 350
              |||||||||||||:        |       ||      ||||||||||||
Db        314 KLIEFLSKFQNDRTE-----DEQFNDEKTYLVKQ----IRDLKRPAQQEA 354


RESULT 8
AAY94247
ID   AAY94247 standard; protein; 337 AA.
XX
AC   AAY94247;
XX
DT   10-AUG-2000  (first entry)
XX
DE   Human calcium binding protein hCBP.
XX
KW   Human; calcium binding protein; cancer; inflammation; CBP;
KW   reproductive disorder; autoimmune disorder; developmental disorder;
KW   seizure disorder; immune disorder; infection.
XX
OS   Homo sapiens.
XX
PN   WO200029580-A1.
XX
PD   25-MAY-2000.
```

```
XX
PF   12-NOV-1999;   99WO-US027027.
XX
PR   13-NOV-1998;   98US-00190965.
XX
PA   (INCY-) INCYTE PHARM INC.
XX
PI   Tang YT,  Guegler KJ,  Corley NC,  Gorgone GA;
XX
DR   WPI; 2000-387793/33.
DR   N-PSDB; AAA27332.
XX
PT   Human hCBP protein, and the nucleic acid encoding it, useful for e.g.
PT   diagnosis, prevention and treatment of cancers, immune, developmental or
PT   reproductive disorders.
XX
PS   Claim 1; Fig 1; 72pp; English.
XX
CC   The present sequence is the human calcium binding protein hCBP. It was
CC   obtained by screening a coronary artery smooth muscle cDNA library, from
CC   which five overlapping nucleic acids were isolated, sequenced and
CC   expressed to give the protein. The protein and the gene encoding it are
CC   useful for the diagnosis and treatment of the following types of
CC   disorder: cancers (such as adenocarcinomas), reproductive disorders (such
CC   as infertility, ovulatory defects, endometriosis, disruptions of the
CC   oestrus and menstrual cycles, polycystic ovary syndrome and ovarian
CC   hyperstimulation), autoimmune disorders (such as benign prostatic
CC   hyperplasia and prostatitis), developmental disorders (such as Cushing's
CC   syndrome, muscular dystrophy and gonadal dysgenesis), hereditary
CC   neuropathies, seizure disorders, immune disorders (such as AIDS,
CC   allergies, anaemia, asthma, atherosclerosis, cholecystitis, Crohn's
CC   disease, diabetes, Graves' disease, multiple sclerosis, psoriasis,
CC   rheumatoid arthritis, scleroderma, Sjogren's syndrome and ulcerative
CC   colitis), and viral, bacterial, fungal, parasitic, protozoal and
CC   helminthic infections
XX
SQ   Sequence 337 AA;

  Query Match             72.9%;  Score 1297.5;  DB 3;  Length 337;
  Best Local Similarity   76.0%;  Pred. No. 6.4e-112;
  Matches  263;  Conservative   32;  Mismatches   38;  Indels   13;  Gaps    3;

Qy          1 MPFPFGKSHKSPADIVKNLKESMAVLEKQDISDKKAEKATEEVSKNLVAMKEILYGTNEK 60
              ||  | ||||:||:||| ||:::|:|||   ||| :||:|||||:| |||||| |||||
Db          4 MPL-FSKSHKNPAEIVKILKDNLAILEKQ---DKKTDKASEEVSKSLQAMKEILCGTNEK 59

Qy         61 EPQTEAVAQLAQELYNSGLLSTLVADLQLIDFEGKKDVAQIFNNILRRQIGTRTPTVEYI 120
              || |||||||||||:|||| ||:|||||||||||||| |||||||||||||:||||||
Db         60 EPPTEAVAQLAQELYSSGLLVTLIADLQLIDFEGKKDVTQIFNNILRRQIGTRSPTVEYI 119

Qy        121 CTQQNILFMLLKGYESPEIALNCGIMLRECIRHEPLAKIILWSEQFYDFFRYVEMSTFDI 180
                 :||||||||||:|:||| |||||||||||||||||:| || |||:|||:|||||
Db        120 SAHPHILFMLLKGYEAPQIALRCGIMLRECIRHEPLAKIILFSNQFRDFFKYVELSTFDI 179

Qy        181 ASDAFATFKDLLTRHKLLSAEFLEQHYDRFFSEYEKLLHSENYVTKRQSLKLLGELLLDR 240
              |||||||||||||||:| |:||||:||  | :||||| |||||||||||||||||:|||
Db        180 ASDAFATFKDLLTRHKVLVADFLEQNYDTIFEDYEKLLQSENYVTKRQSLKLLGELILDR 239

Qy        241 HNFTIMTKYISKPENLKLMMNLLRDKSRNIQFEAFHVFKVFVANPNKTQPILDILLKNQA 300
              ||| |||||||||||||||||||| ||||||||||||||:|:|||||::||||||
Db        240 HNFAIMTKYISKPENLKLMMNLLRDKSPNIQFEAFHVFKVFVASPHKTQPIVEILLKNQP 299

Qy        301 KLIEFLSKFQNDRTDCMSSSVPTTNSRVDLRVKPRTRGIRDLKRPA 346
              ||||||| || :|||       : :       : |||||: |
Db        300 KLIEFLSSFQKERTD---------DEQFADEKNYLIKQIRDLKKTA 336


RESULT 9
AAB82090
ID   AAB82090 standard; protein; 337 AA.
XX
AC   AAB82090;
XX
DT   26-JUN-2001  (first entry)
XX
DE   Human Acute Neuronal Induced Calcium Binding Protein, ANIC-BP.
XX
KW   Human; cerebroprotective; neuroprotective; vulnerary; vaccine;
KW   gene therapy; Acute Neuronal Induced Calcium Binding Protein; ANIC-BP;
KW   stroke; acute head trauma; multiple sclerosis; spinal cord injury.
XX
OS   Homo sapiens.
XX
PN   WO200123552-A1.
XX
PD   05-APR-2001.
XX
PF   18-SEP-2000; 2000WO-EP009132.
XX
PR   24-SEP-1999;   99EP-00118848.
XX
PA   (MERE ) MERCK PATENT GMBH.
XX
PI   Den Daas I,  Duecker K;
XX
DR   WPI; 2001-308142/32.
DR   N-PSDB; AAF86462.
XX
PT   Novel human acute neuronal induced calcium binding polypeptide, and
PT   polynucleotides encoding them useful for diagnosing or treating stroke,
PT   acute head trauma, multiple sclerosis and spinal cord injury.
XX
PS   Claim 1; Page 41-42; 45pp; English.
XX
CC   The present sequence is the protein sequence for human Acute Neuronal
CC   Induced Calcium Binding Protein (ANIC-BP). ANIC-BP coding sequence and
CC   protein are useful for treating stroke, acute head trauma, multiple
CC   sclerosis and spinal cord injury. ANIC-BP coding sequence and protein are
CC   also useful as vaccines for inducing an immunological response in a
CC   mammal
XX
SQ   Sequence 337 AA;

  Query Match             72.9%;  Score 1297.5;  DB 4;  Length 337;
  Best Local Similarity   76.0%;  Pred. No. 6.4e-112;
  Matches  263;  Conservative   32;  Mismatches   38;  Indels   13;  Gaps    3;

Qy          1 MPFPFGKSHKSPADIVKNLKESMAVLEKQDISDKKAEKATEEVSKNLVAMKEILYGTNEK 60
              ||  | ||||:||:||| ||:::|:|||   ||| :||:|||||:| |||||| |||||
Db          4 MPL-FSKSHKNPAEIVKILKDNLAILEKQ---DKKTDKASEEVSKSLQAMKEILCGTNEK 59

Qy         61 EPQTEAVAQLAQELYNSGLLSTLVADLQLIDFEGKKDVAQIFNNILRRQIGTRTPTVEYI 120
              || |||||||||||:|||| ||:|||||||||||||| |||||||||||||:||||||
Db         60 EPPTEAVAQLAQELYSSGLLVTLIADLQLIDFEGKKDVTQIFNNILRRQIGTRSPTVEYI 119

Qy        121 CTQQNILFMLLKGYESPEIALNCGIMLRECIRHEPLAKIILWSEQFYDFFRYVEMSTFDI 180
                 :||||||||||:|:||| |||||||||||||||||:| || |||:|||:|||||
Db        120 SAHPHILFMLLKGYEAPQIALRCGIMLRECIRHEPLAKIILFSNQFRDFFKYVELSTFDI 179

Qy        181 ASDAFATFKDLLTRHKLLSAEFLEQHYDRFFSEYEKLLHSENYVTKRQSLKLLGELLLDR 240
              |||||||||||||||:| |:||||:||  | :||||| |||||||||||||||||:|||
Db        180 ASDAFATFKDLLTRHKVLVADFLEQNYDTIFEDYEKLLQSENYVTKRQSLKLLGELILDR 239

Qy        241 HNFTIMTKYISKPENLKLMMNLLRDKSRNIQFEAFHVFKVFVANPNKTQPILDILLKNQA 300
              ||| |||||||||||||||||||| ||||||||||||||:|:|||||::||||||
Db        240 HNFAIMTKYISKPENLKLMMNLLRDKSPNIQFEAFHVFKVFVASPHKTQPIVEILLKNQP 299

Qy        301 KLIEFLSKFQNDRTDCMSSSVPTTNSRVDLRVKPRTRGIRDLKRPA 346
              ||||||| || :|||       : :       : |||||: |
Db        300 KLIEFLSSFQKERTD---------DEQFADEKNYLIKQIRDLKKTA 336


RESULT 10
AAM39078
ID   AAM39078 standard; protein; 337 AA.
XX
```

```
AC   AAM39078;
XX
DT   22-OCT-2001  (first entry)
XX
DE   Human polypeptide SEQ ID NO 2223.
XX
KW   Human; nootropic; immunosuppressant; cytostatic; gene therapy; cancer;
KW   peripheral nervous system; neuropathy; central nervous system; CNS;
KW   Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;
KW   amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;
KW   chemokinetic; thrombolytic; drug screening; arthritis; inflammation;
KW   leukaemia.
XX
OS   Homo sapiens.
XX
PN   WO200153312-A1.
XX
PD   26-JUL-2001.
XX
PF   26-DEC-2000; 2000WO-US034263.
XX
PR   23-DEC-1999;   99US-00471275.
PR   21-JAN-2000; 2000US-00488725.
PR   25-APR-2000; 2000US-00552317.
PR   20-JUN-2000; 2000US-00598042.
PR   19-JUL-2000; 2000US-00620312.
PR   03-AUG-2000; 2000US-00653450.
PR   14-SEP-2000; 2000US-00662191.
PR   19-OCT-2000; 2000US-00693036.
PR   29-NOV-2000; 2000US-00727344.
XX
PA   (HYSE-) HYSEQ INC.
XX
PI   Tang YT,  Liu C,  Asundi V,  Chen R,  Ma Y,  Qian XB,  Ren F,  Wang D;
PI   Wang J,  Wang Z,  Wehrman T,  Xu C,  Xue AJ,  Yang Y,  Zhang J,  Zhao QA;
PI   Zhou P,  Goodrich R,  Drmanac RT;
XX
DR   WPI; 2001-442253/47.
DR   N-PSDB; AAI58234.
XX
PT   Novel nucleic acids and polypeptides, useful for treating disorders such
PT   as central nervous system injuries.
XX
PS   Example 4; SEQ ID NO 2223; 10078pp; English.
XX
CC   The invention relates to human nucleic acids (AAI57798-AAI61369) and the
CC   encoded polypeptides (AAM38642-AAM42213) with nootropic,
CC   immunosuppressant and cytostatic activity. The polynucleotides are useful
CC   in gene therapy. A composition containing a polypeptide or polynucleotide
CC   of the invention may be used to treat diseases of the peripheral nervous
CC   system, such as peripheral nervous injuries, peripheral neuropathy and
CC   localised neuropathies and central nervous system diseases, such as
CC   Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic
CC   lateral sclerosis, and Shy-Drager Syndrome. Other uses include the
CC   utilisation of the activities such as: Immune system suppression,
CC   Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic
CC   and thrombolytic activity, cancer diagnosis and therapy, drug screening,
CC   assays for receptor activity, arthritis and inflammation, leukaemias and
CC   C.N.S disorders. Note: The sequence data for this patent did not form
CC   part of the printed specification
XX
SQ   Sequence 337 AA;

  Query Match             72.9%;  Score 1297.5;  DB 4;  Length 337;
  Best Local Similarity   76.0%;  Pred. No. 6.4e-112;
  Matches  263;  Conservative   32;  Mismatches   38;  Indels   13;  Gaps    3;

Qy          1 MPFPFGKSHKSPADIVKNLKESMAVLEKQDISDKKAEKATEEVSKNLVAMKEILYGTNEK 60
              ||  | ||||:||:||| ||:::|:||||   ||| :||:|||||:| |||||| |||||
Db          4 MPL-FSKSHKNPAEIVKILKDNLAILEKQ---DKKTDKASEEVSKSLQAMKEILCGTNEK 59

Qy         61 EPQTEAVAQLAQELYNSGLLSTLVADLQLIDFEGKKDVAQIFNNILRRQIGTRTPTVEYI 120
              || ||||||||||||:|||| ||:|||||||||||||| ||||||||||||||:||||||

Db         60 EPPTEAVAQLAQELYSSGLLVTLIADLQLIDFEGKKDVTQIFNNILRRQIGTRSPTVEYI 119

Qy        121 CTQQNILFMLLKGYESPEIALNCGIMLRECIRHEPLAKIILWSEQFYDFFRYVEMSTFDI 180
                :|||||||||:|:|||| ||||||||||||||||||:| || ||||:|||:|||||
Db        120 SAHPHILFMLLKGYEAPQIALRCGIMLRECIRHEPLAKIILFSNQFRDFFKYVELSTFDI 179

Qy        181 ASDAFATFKDLLTRHKLLSAEFLEQHYDRFFSEYEKLLHSENYVTKRQSLKLLGELLLDR 240
              |||||||||||||||:| |:||||:||  | :||||| ||||||||||||||||:|||
Db        180 ASDAFATFKDLLTRHKVLVADFLEQNYDTIFEDYEKLLQSENYVTKRQSLKLLGELILDR 239

Qy        241 HNFTIMTKYISKPENLKLMMNLLRDKSRNIQFEAFHVFKVFVANPNKTQPILDILLKNQA 300
              ||| ||||||||||||||||||||| ||||||||||||||:|:||||::||||||
Db        240 HNFAIMTKYISKPENLKLMMNLLRDKSPNIQFEAFHVFKVFVASPHKTQPIVEILLKNQP 299

Qy        301 KLIEFLSKFQNDRTDCMSSSVPTTNSRVDLRVKPRTRGIRDLKRPA 346
              ||||||| || :|||         : :         : |||||: |
Db        300 KLIEFLSSFQKERTD---------DEQFADEKNYLIKQIRDLKKTA 336


RESULT 11
AAB94139
ID   AAB94139 standard; protein; 289 AA.
XX
AC   AAB94139;
XX
DT   26-JUN-2001  (first entry)
XX
DE   Human protein sequence SEQ ID NO:14408.
XX
KW   Human; primer; detection; diagnosis; antisense therapy; gene therapy.
XX
OS   Homo sapiens.
XX
PN   EP1074617-A2.
XX
PD   07-FEB-2001.
XX
PF   28-JUL-2000; 2000EP-00116126.
XX
PR   29-JUL-1999;   99JP-00248036.
PR   27-AUG-1999;   99JP-00300253.
PR   11-JAN-2000; 2000JP-00118776.
PR   02-MAY-2000; 2000JP-00183767.
PR   09-JUN-2000; 2000JP-00241899.
XX
PA   (HELI-) HELIX RES INST.
XX
PI   Ota T,  Isogai T,  Nishikawa T,  Hayashi K,  Saito K,  Yamamoto J;
PI   Ishii S,  Sugiyama T,  Wakamatsu A,  Nagai K,  Otsuki T;
XX
DR   WPI; 2001-318749/34.
XX
PT   Primer sets for synthesizing polynucleotides, particularly the 5602 full-
PT   length cDNAs defined in the specification, and for the detection and/or
PT   diagnosis of the abnormality of the proteins encoded by the full-length
PT   cDNAs.
XX
PS   Claim 8; SEQ ID NO 14408; 2537pp + Sequence Listing; English.
XX
CC   The present invention describes primer sets for synthesising 5602 full-
CC   length cDNAs defined in the specification. Where a primer set comprises:
CC   (a) an oligo-dT primer and an oligonucleotide complementary to the
CC   complementary strand of a polynucleotide which comprises one of the 5602
CC   nucleotide sequences defined in the specification, where the
CC   oligonucleotide comprises at least 15 nucleotides; or (b) a combination
CC   of an oligonucleotide comprising a sequence complementary to the
CC   complementary strand of a polynucleotide which comprises a 5'-end
CC   sequence and an oligonucleotide comprising a sequence complementary to a
CC   polynucleotide which comprises a 3'-end sequence, where the
CC   oligonucleotide comprises at least 15 nucleotides and the combination of
CC   the 5'-end sequence/3'-end sequence is selected from those defined in the
CC   specification. The primer sets can be used in antisense therapy and in
CC   gene therapy. The primers are useful for synthesising polynucleotides,
```

CC particularly full-length cDNAs. The primers are also useful for the
CC detection and/or diagnosis of the abnormality of the proteins encoded by
CC the full-length cDNAs. The primers allow obtaining of the full-length
CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
CC AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to AAB95893
CC represent human amino acid sequences; and AAH13629 to AAH13632 represent
CC oligonucleotides, all of which are used in the exemplification of the
CC present invention
XX
SQ Sequence 289 AA;

Query Match 64.8%; Score 1153.5; DB 4; Length 289;
Best Local Similarity 77.8%; Pred. No. 1.3e-98;
Matches 231; Conservative 23; Mismatches 34; Indels 9; Gaps 1;

```
Qy         50 MKEILYGTNEKEPQTEAVAQLAQELYNSGLLSTLVADLQLIDFEGKKDVAQIFNNILRRQ 109
              ||||| ||||||| |||||||||||||:|||| ||:||||||||||||| ||||||||||
Db          1 MKEILCGTNEKEPPTEAVAQLAQELYSSGLLVTLIADLQLIDFEGKKDVTQIFNNILRRQ 60

Qy        110 IGTRTPTVEYICTQQNILFMLLKGYESPEIALNCGIMLRECIRHEPLAKIILWSEQFYDF 169
              ||||:|||||    :|||||||||:|:||| |||||||||||||| ||||:| || ||
Db         61 IGTRSPTVEYISAHPHILFMLLKGYEAPQIALRCGIMLRECIRHEPLVKIILFSNQFRDF 120

Qy        170 FRYVEMSTFDIASDAFATFKDLLTRHKLLSAEFLEQHYDRFFSEYEKLLHSENYVTKRQS 229
              |:|||:|||||||||||||||||||:| |:||||:||  | :||||| ||||||||||
Db        121 FKYVELSTFDIASDAFATFKDLLTRHKVLVADFLEQNYDTIFEDYEKLLQSENYVTKRQS 180

Qy        230 LKLLGELLLDRHNFTIMTKYISKPENLKLMMNLLRDKSRNIQFEAFHVFKVFVANPNKTQ 289
              |||||||:|||||| |||||||||||||||||||||||| ||||||||||||||:|:|||
Db        181 LKLLGELILDRHNFAIMTKYISKPENLKLMMNLLRDKSPNIQFEAFHVFKVFVASPHKTQ 240

Qy        290 PILDILLKNQAKLIEFLSKFQNDRTDCMSSSVPTTNSRVDLRVKPRTRGIRDLKRPA 346
              ||::|||||| ||||||| || :|||         : :         : |||||: |
Db        241 PIVEILLKNQPKLIEFLSSFQKERTD---------DEQFADEKNYLIKQIRDLKKTA 288
```

RESULT 12
ABB60392
ID ABB60392 standard; protein; 339 AA.
XX
AC ABB60392;
XX
DT 26-MAR-2002 (first entry)
XX
DE Drosophila melanogaster polypeptide SEQ ID NO 7968.
XX
KW Drosophila; developmental biology; cell signalling; insecticide;
KW pharmaceutical.
XX
OS Drosophila melanogaster.
XX
PN WO200171042-A2.
XX
PD 27-SEP-2001.
XX
PF 23-MAR-2001; 2001WO-US009231.
XX
PR 23-MAR-2000; 2000US-0191637P.
PR 11-JUL-2000; 2000US-00614150.
XX
PA (PEKE) PE CORP NY.
XX
PI Venter JC, Adams M, Li PWD, Myers EW;
XX
DR WPI; 2001-656860/75.
DR N-PSDB; ABL04495.
XX
PT New isolated nucleic acid detection reagent for detecting 1000 or more
PT genes from Drosophila and for elucidating cell signaling and cell-cell
PT interactions.
XX
PS Disclosure; SEQ ID NO 7968; 21pp + Sequence Listing; English.
XX

CC The invention relates to an isolated nucleic acid detection reagent
CC capable of detecting 1000 or more genes from Drosophila. The invention is
CC useful in developmental biology and in elucidating cell signalling and
CC cell-cell interactions in higher eukaryotes for the development of
CC insecticides, therapeutics and pharmaceutical drugs. The invention
CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC sequences (ABL01840-ABL16175) and the encoded proteins (ABB57737-
CC ABB72072). The sequence data for this patent did not form part of the
CC printed specification, but was obtained in electronic format directly
CC from WIPO at ftp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 339 AA;

Query Match 63.1%; Score 1123; DB 4; Length 339;
Best Local Similarity 69.3%; Pred. No. 1.1e-95;
Matches 221; Conservative 50; Mismatches 40; Indels 8; Gaps 5;

```
Qy          1 MPFPFGKSHKSPADIVKNLKESMAVLEKQDISDKKAEKATEEVSKNLVAMKEILYGTNEK 60
              ||  |||| ||| ::||:|||::  ||      |:| ||| |:|||||::| :|||:::
Db          1 MPL-FGKSQKSPVELVKSLKEAINALE---AGDRKVEKAQEDVSKNLVSIKNMLYGSSDA 56

Qy         61 EPQTE-AVAQLAQELYNSGLLSTLVADLQLIDFEGKKDVAQIFNNILRRQIGTRTPTVEY 119
              ||  :  ||||:|||||| ||  |: :|  ||||||| || ||||:|||||||:|||||
Db         57 EPPADYVVAQLSQELYNSNLLLLLIQNLHRIDFEGKKHVALIFNNVLRRQIGTRSPTVEY 116

Qy        120 ICTQQNILFMLLKGYES--PEIALNCGIMLRECIRHEPLAKIILWSEQFYDFFRYVEMST 177
              |||:  ||| |: |||   |||||| | ||||| |:| ||||:| |::|: ||||||:||
Db        117 ICTKPEILFTLMAGYEDAHPEIALNSGTMLRECARYEALAKIMLHSDEFFKFFRYVEVST 176

Qy        178 FDIASDAFATFKDLLTRHKLLSAEFLEQHYDRFFSE-YEKLLHSENYVTKRQSLKLLGEL 236
              |||||||||:|||:|||||||| |||: :||:|||: |::||:|||||:||||||||||
Db        177 FDIASDAFSTFKELLTRHKLLCAEFLDANYDKFFSQHYQRLLNSENYVTRRQSLKLLGEL 236

Qy        237 LLDRHNFTIMTKYISKPENLKLMMNLLRDKSRNIQFEAFHVFKVFVANPNKTQPILDILL 296
              |||||||:||:|||:|||||||||:|::|||||||||||||||||||||| :|||||||
Db        237 LLDRHNFTVMTRYISEPENLKLMMNMLKEKSRNIQFEAFHVFKVFVANPNKPKPILDILL 296

Qy        297 KNQAKLIEFLSKFQNDRTD 315
              :|| ||::||: |  ||::
Db        297 RNQTKLVDFLTNFHTDRSE 315
```

RESULT 13
AAY94249
ID AAY94249 standard; protein; 339 AA.
XX
AC AAY94249;
XX
DT 10-AUG-2000 (first entry)
XX
DE Drosophila calcium binding protein DMO25.
XX
KW Drosophila; calcium binding protein; cancer; inflammation; DMO25; CBP;
KW reproductive disorder; autoimmune disorder; developmental disorder;
KW seizure disorder; immune disorder; infection.
XX
OS Drosophila melanogaster.
XX
PN WO200029580-A1.
XX
PD 25-MAY-2000.
XX
PF 12-NOV-1999; 99WO-US027027.
XX
PR 13-NOV-1998; 98US-00190965.
XX
PA (INCY-) INCYTE PHARM INC.
XX
PI Tang YT, Guegler KJ, Corley NC, Gorgone GA;
XX
DR WPI; 2000-387793/33.
XX
PT Human hCBP protein, and the nucleic acid encoding it, useful for e.g.

PT diagnosis, prevention and treatment of cancers, immune, developmental or
PT reproductive disorders.
XX
PS Disclosure; Page 67-68; 72pp; English.
XX
CC The present sequence is the Drosophila calcium binding protein DMO25. It
CC was used in a sequence alignment to identify human calcium binding
CC protein hCBP. The hCBP protein and the gene encoding it are useful for
CC the diagnosis and treatment of the following types of disorder: cancers
CC (such as adenocarcinomas), reproductive disorders (such as infertility,
CC ovulatory defects, endometriosis, disruptions of the oestrus and
CC menstrual cycles, polycystic ovary syndrome and ovarian
CC hyperstimulation), autoimmune disorders (such as benign prostatic
CC hyperplasia and prostatitis), developmental disorders (such as Cushing's
CC syndrome, muscular dystrophy and gonadal dysgenesis), hereditary
CC neuropathies, seizure disorders, immune disorders (such as AIDS,
CC allergies, anaemia, asthma, atherosclerosis, cholecystitis, Crohn's
CC disease, diabetes, Graves' disease, multiple sclerosis, psoriasis,
CC rheumatoid arthritis, scleroderma, Sjogren's syndrome and ulcerative
CC colitis), and viral, bacterial, fungal, parasitic, protozoal and
CC helminthic infections
XX
SQ Sequence 339 AA;

Query Match 62.7%; Score 1117; DB 3; Length 339;
Best Local Similarity 69.0%; Pred. No. 4.1e-95;
Matches 220; Conservative 51; Mismatches 40; Indels 8; Gaps 5;

```
Qy          1 MPFPPFGKSHKSPADIVKNLKESMAVLEKQDISDKKAEKATEEVSKNLVAMKEILYGTNEK 60
              ||  |||| ||| ::||:|||::  ||     |:| ||| |:||||||::| :|:|:::
Db          1 MPL-FGKSQKSPVELVKSLKEAINALE---AGDRKVEKAQEDVSKNLVSIKNMLHGSSDA 56

Qy         61 EPQTE-AVAQLAQELYNSGLLSTLVADLQLIDFEGKKDVAQIFNNILRRQIGTRTPTVEY 119
              ||  :  ||||:|||||| ||  |: :|  ||||||| || ||||:||||||||:|||||
Db         57 EPPADYVVAQLSQELYNSNLLLLLIQNLHRIDFEGKKHVALIFNNLLRRQIGTRSPTVEY 116

Qy        120 ICTQQNILFMLLKGYES--PEIALNCGIMLRECIRHEPLAKIILWSEQFYDFFRYVEMST 177
              |||:  ||| |: |||   |||||| | ||||| |:| ||||:| |::|: ||||||:||
Db        117 ICTKPEILFTLMAGYEDAHPEIALNSGTMLRECARYEALAKIMLHSDEFFKFFRYVEVST 176

Qy        178 FDIASDAFATFKDLLTRHKLLSAEFLEQHYDRFFSE-YEKLLHSENYVTKRQSLKLLGEL 236
              ||||||||:|||:|||||||| |||||: :||:|||: |::||:||||||:|||||||||
Db        177 FDIASDAFSTFKELLTRHKLLCAEFLDANYDKFFSQHYQRLLNSENYVTRRQSLKLLGEL 236

Qy        237 LLDRHNFTIMTKYISKPENLKLMMNLLRDKSRNIQFEAFHVFKVFVANPNKTQPILDILL 296
              ||||||||:||:|||:|||||||||:|::|||||||||||||||||||||| :|||||||
Db        237 LLDRHNFTVMTRYISEPENLKLMMNMLKEKSRNIQFEAFHVFKVFVANPNKPKPILDILL 296

Qy        297 KNQAKLIEFLSKFQNDRTD 315
              :|| ||::||: |  ||::
Db        297 RNQTKLVDFLTNFHTDRSE 315
```

RESULT 14
AAY94250
ID AAY94250 standard; protein; 377 AA.
XX
AC AAY94250;
XX
DT 10-AUG-2000 (first entry)
XX
DE C. elegans yeast-like calcium binding protein.
XX
KW Calcium binding protein; cancer; inflammation; yeast-like CBP; CBP;
KW reproductive disorder; autoimmune disorder; developmental disorder;
KW seizure disorder; immune disorder; infection.
XX
OS Caenorhabditis elegans.
XX
PN WO200029580-A1.
XX
PD 25-MAY-2000.
XX
PF 12-NOV-1999; 99WO-US027027.
XX
PR 13-NOV-1998; 98US-00190965.
XX
PA (INCY-) INCYTE PHARM INC.
XX
PI Tang YT, Guegler KJ, Corley NC, Gorgone GA;
XX
DR WPI; 2000-387793/33.
XX
PT Human hCBP protein, and the nucleic acid encoding it, useful for e.g.
PT diagnosis, prevention and treatment of cancers, immune, developmental or
PT reproductive disorders.
XX
PS Disclosure; Page 68-69; 72pp; English.
XX
CC The present sequence is the C. elegans yeast-like CBP. It was used in a
CC sequence alignment to identify human calcium binding protein hCBP. The
CC hCBP protein and the gene encoding it are useful for the diagnosis and
CC treatment of the following types of disorder: cancers (such as
CC adenocarcinomas), reproductive disorders (such as infertility, ovulatory
CC defects, endometriosis, disruptions of the oestrus and menstrual cycles,
CC polycystic ovary syndrome and ovarian hyperstimulation), autoimmune
CC disorders (such as benign prostatic hyperplasia and prostatitis),
CC developmental disorders (such as Cushing's syndrome, muscular dystrophy
CC and gonadal dysgenesis), hereditary neuropathies, seizure disorders,
CC immune disorders (such as AIDS, allergies, anaemia, asthma,
CC atherosclerosis, cholecystitis, Crohn's disease, diabetes, Graves'
CC disease, multiple sclerosis, psoriasis, rheumatoid arthritis,
CC scleroderma, Sjogren's syndrome and ulcerative colitis), and viral,
CC bacterial, fungal, parasitic, protozoal and helminthic infections
XX
SQ Sequence 377 AA;

Query Match 59.2%; Score 1054.5; DB 3; Length 377;
Best Local Similarity 62.5%; Pred. No. 3.1e-89;
Matches 205; Conservative 49; Mismatches 61; Indels 13; Gaps 2;

```
Qy          1 MPFPPFGKSHKSPADIVKNLKESMAVLEK----------QDISDKKAEKATEEVSKNLVAM 50
              ||  |||||||||||:|| |:| : :|:|              |||| :|| :|||||:  :
Db          1 MPLLFGKSHKSPADVVKTLREVLTILDKLPPPKLDKDGNIQSDKKYDKALDEVSKNVAMI 60

Qy         51 KEILYGTNEKEPQTE---AVAQLAQELYNSGLLSTLVADLQLIDFEGKKDVAQIFNNILR 107
              |  :|| :  || :|    |||||||:||: :|  |:  |   :|| |||| |||||:||
Db         61 KSFIYGNDSAEPSSEHVVQVAQLAQEVYNANILPMLIKMLPKFEFECKKDVGQIFNNLLR 120

Qy        108 RQIGTRTPTVEYICTQQNILFMLLKGYESPEIALNCGIMLRECIRHEPLAKIILWSEQFY 167
              ||||||:|||||:  :  ||  |::||  |:||| ||:|||| ||||: ||||||:|: ||
Db        121 RQIGTRSPTVEYLGARPEILIQLVQGYSVPDIALTCGLMLRESIRHDHLAKIILYSDVFY 180

Qy        168 DFFRYVEMSTFDIASDAFATFKDLLTRHKLLSAEFLEQHYDRFFSEYEKLLHSENYVTKR 227
               || ||:   |||:||||:|||:| |||| : ||||: :|| ||::|: ||:|:||||:|
Db        181 TFFLYVQSEVFDISSDAFSTFKELTTRHKAIIAEFLDSNYDTFFAQYQNLLNSKNYVTRR 240

Qy        228 QSLKLLGELLLDRHNFTIMTKYISKPENLKLMMNLLRDKSRNIQFEAFHVFKVFVANPNK 287
              ||||||||||||||| |||||| |:||:||| ||||||||||:||||||||||||||||
Db        241 QSLKLLGELLLDRHNFNTMTKYISNPDNLRLMMELLRDKSRNIQYEAFHVFKVFVANPNK 300

Qy        288 TQPILDILLKNQAKLIEFLSKFQNDRTD 315
               :|| ||| :|: ||:||||:| |||||
Db        301 PKPISDILNRNREKLVEFLSEFHNDRTD 328
```

RESULT 15
AAM40864
ID AAM40864 standard; protein; 237 AA.
XX
AC AAM40864;
XX
DT 22-OCT-2001 (first entry)
XX
DE Human polypeptide SEQ ID NO 5795.
XX

KW Human; nootropic; immunosuppressant; cytostatic; gene therapy; cancer;
KW peripheral nervous system; neuropathy; central nervous system; CNS;
KW Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;
KW amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;
KW chemokinetic; thrombolytic; drug screening; arthritis; inflammation;
KW leukaemia.
XX
OS Homo sapiens.
XX
PN WO200153312-A1.
XX
PD 26-JUL-2001.
XX
PF 26-DEC-2000; 2000WO-US034263.
XX
PR 23-DEC-1999; 99US-00471275.
PR 21-JAN-2000; 2000US-00488725.
PR 25-APR-2000; 2000US-00552317.
PR 20-JUN-2000; 2000US-00598042.
PR 19-JUL-2000; 2000US-00620312.
PR 03-AUG-2000; 2000US-00653450.
PR 14-SEP-2000; 2000US-00662191.
PR 19-OCT-2000; 2000US-00693036.
PR 29-NOV-2000; 2000US-00727344.
XX
PA (HYSE-) HYSEQ INC.
XX
PI Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;
PI Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J, Zhao QA;
PI Zhou P, Goodrich R, Drmanac RT;
XX
DR WPI; 2001-442253/47.
DR N-PSDB; AAI60020.
XX
PT Novel nucleic acids and polypeptides, useful for treating disorders such
PT as central nervous system injuries.
XX
PS Example 2; SEQ ID NO 5795; 10078pp; English.
XX
CC The invention relates to human nucleic acids (AAI57798-AAI61369) and the
CC encoded polypeptides (AAM38642-AAM42213) with nootropic,
CC immunosuppressant and cytostatic activity. The polynucleotides are useful
CC in gene therapy. A composition containing a polypeptide or polynucleotide
CC of the invention may be used to treat diseases of the peripheral nervous
CC system, such as peripheral nervous injuries, peripheral neuropathy and
CC localised neuropathies and central nervous system diseases, such as
CC Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic
CC lateral sclerosis, and Shy-Drager Syndrome. Other uses include the
CC utilisation of the activities such as: Immune system suppression,
CC Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic
CC and thrombolytic activity, cancer diagnosis and therapy, drug screening,
CC assays for receptor activity, arthritis and inflammation, leukaemias and
CC C.N.S disorders. Note: The sequence data for this patent did not form
CC part of the printed specification
XX
SQ Sequence 237 AA;

Query Match 49.9%; Score 888.5; DB 4; Length 237;
Best Local Similarity 75.1%; Pred. No. 4.5e-74;
Matches 178; Conservative 20; Mismatches 34; Indels 5; Gaps 2;

```
Qy        112 TRTPTVEYICTQQNILFMLLKGYESPEIALNCGIMLRECIRHEPLAKIILWSEQFYDFFR 171
              ||:||||||    :|||||||||||:|:||| |||||||||||||||||||:| || |||:
Db          2 TRSPTVEYISAHPHILFMLLKGYEAPQIALRCGIMLRECIRHEPLAKIILFSNQFRDFFK 61

Qy        172 YVEMSTFDIASDAFATFKDLLTRHKLLSAEFLEQHYDRFFSEYEKLLHSENYVTKRQSLK 231
              |||:|||||||||||||||||||:| |:||||:||  | :||||| ||||||||||||
Db         62 YVELSTFDIASDAFATFKDLLTRHKVLVADFLEQNYDTIFEDYEKLLQSENYVTKRQSLK 121

Qy        232 LLGELLLDRHNFTIMTKYISKPENLKLMMNLLRDKSRNIQFEAFHVFKVFVANPNKTQPI 291
              |||||:|||||| ||||||||||||||||||||| |||||||||||||:|:||||||
Db        122 LLGELILDRHNFAIMTKYISKPENLKLMMNLLRDKSPNIQFEAFHVFKVFVASPHKTQPI 181
```

```
Qy        292 LDILLKNQAKLIEFLSKFQNDRTDCMSSSVPTTNSRV----DLRVKPRTRGIRDLKR 344
              ::||||| ||||||| || :|||         |  :    ||:     | :|| ||
Db        182 VEILLKNQPKLIEFLSSFQKERTD-DEQFADEKNYLIKQIRDLKKTAPXRALRDSKR 237
```

Search completed: October 22, 2004, 09:09:07
Job time : 162 secs

GenCore version 5.1.6

OM protein - protein search, using sw model

Run on: October 22, 2004, 09:03:58 ; Search time 39 Seconds
(without alignments)
863.484 Million cell updates/sec

Title: US-10-089-688-2
Perfect score: 1781
Sequence: 1 MPFPPFGKSHKSPADIVKNLK.....RVKPRTRGIRDLKRPAQQEA 350

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283416 seqs, 96216763 residues

Total number of hits satisfying chosen parameters: 283416

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR_79:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	1614.5	90.7	341	2	I57997	hypothetical calci
2	1054.5	59.2	377	2	T16651	hypothetical prote
3	1006	56.5	338	2	T27129	hypothetical prote
4	766	43.0	329	2	T50117	mo25 homolog [impo
5	666.5	37.4	305	2	G71441	hypothetical prote
6	614	34.5	348	2	B84448	hypothetical prote
7	501	28.1	399	2	S34681	hypothetical prote
8	137.5	7.7	339	2	T33477	hypothetical prote
9	122	6.9	602	2	T03057	hypothetical prote
10	117	6.6	365	2	D97848	hypothetical prote
11	115	6.5	2819	2	A90551	conserved hypothet
12	113.5	6.4	1939	2	T18372	repeat organellar
13	112.5	6.3	677	2	H64574	DNA topoisomerase
14	111	6.2	4981	2	T18489	hypothetical prote
15	109	6.1	2166	2	G70163	hypothetical prote
16	108.5	6.1	719	2	A81358	hypothetical prote
17	108	6.1	2401	2	T28676	rhoptry protein -
18	107.5	6.0	1050	2	A89769	hypothetical prote
19	106.5	6.0	348	2	D70195	hypothetical prote
20	106	6.0	824	2	T20477	hypothetical prote
21	105	5.9	809	1	JQ0032	anthrax toxin leth
22	104.5	5.9	617	2	D96978	hypothetical prote
23	104.5	5.9	1156	2	B70356	chromosome assembl
24	104	5.8	299	1	F70330	probable arsenical
25	104	5.8	592	2	T29402	hypothetical prote
26	104	5.8	4131	2	T21085	hypothetical prote
27	103.5	5.8	415	2	A41253	kinase-related tra
28	103.5	5.8	1189	2	A54817	ATPase ScII, chrom
29	103.5	5.8	1642	2	T08880	NMDA receptor-bind
30	103.5	5.8	2473	1	S38040	1-phosphatidylinos
31	103	5.8	256	2	B81394	hypothetical prote
32	103	5.8	355	2	C64644	integrase/recombin
33	103	5.8	799	2	T00331	hypothetical prote
34	103	5.8	959	2	T00246	DNA polymerase V -
35	102.5	5.8	472	2	T19375	hypothetical prote
36	102.5	5.8	600	2	T09676	probable ATP-depen
37	102.5	5.8	1300	2	T18364	ro-3 protein - Neu
38	102.5	5.8	1666	2	T38393	clathrin heavy cha
39	102	5.7	717	2	T27066	hypothetical prote
40	102	5.7	717	2	T27067	hypothetical prote
41	102	5.7	878	2	T23537	hypothetical prote
42	102	5.7	1026	2	T03108	DNA-directed DNA p
43	101.5	5.7	1170	2	A56157	chromosome segrega
44	101	5.7	1285	2	B72420	hypothetical prote
45	101	5.7	1317	2	A54831	nuclear pore compl

ALIGNMENTS

RESULT 1
I57997
hypothetical calcium-binding protein - mouse
C;Species: Mus sp. (mouse)
C;Date: 02-Aug-1996 #sequence_revision 02-Aug-1996 #text_change 19-May-2000
C;Accession: I57997
R;Miyamoto, H.; Matsushiro, A.; Nozaki, M.
Mol. Reprod. Dev. 34, 1-7, 1993
A;Title: Molecular cloning of a novel mRNA sequence expressed in cleavage stage mouse emb
A;Reference number: I57997; MUID:93119656; PMID:8418809
A;Accession: I57997
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 1-341 <RES>
A;Cross-references: GB:S51858; NID:g262933; PIDN:AAB24801.1; PID:g262934
C;Superfamily: Saccharomyces hypothetical protein YKL189w
C;Keywords: calcium binding

Query Match 90.7%; Score 1614.5; DB 2; Length 341;
Best Local Similarity 92.9%; Pred. No. 7.2e-99;
Matches 325; Conservative 2; Mismatches 14; Indels 9; Gaps 2;

```
Qy          1 MPFPPFGKSHKSPADIVKNLKESMAVLEKQDISDKKAEKATEEVSKNLVAMKEILYGTNEK 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MPFPPFGKSHKSPADIVKNLKESMAVLEKQDISDKKAEKATEEVSKNLVAMKEILYGTNEK 60

Qy         61 EPQTEAVAQLAQELYNSGLLSTLVADLQLIDFEGKKDVAQIFNNILRRQIGTRTPTVEYI 120
              |||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||||
Db         61 EPQTEAVAQLAQELYNSGLLGTLVADLQLIDFEGKKDVAQIFNNILRRQIGTRTPTVEYI 120

Qy        121 CTQQNILFMLLKGYESPEIALNCGIMLRECIRHEPLAKIILWSEQFYDFFRYVEMSTFDI 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CTQQNILFMLLKGYESPEIALNCGIMLRECIRHEPLAKIILWSEQFYDFFRYVEMSTFDI 180

Qy        181 ASDAFATFKDLLTRHKLLSAEFLEQHYDRFFSEYEKLLHSENYVTKRQSLKLLGELLLDR 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 ASDAFATFKDLLTRHKLLSAEFLEQHYDRFFSEYEKLLHSENYVTKRQSLKLLGELLLDR 240

Qy        241 HNFTIMTKYISKPENLKLMMNLLRDKSRNIQFEAFHVFKVFVANPNKTQPILDILLKNQA 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 HNFTIMTKYISKPENLKLMMNLLRDKSRNIQFEAFHVFKVFVANPNKTQPILDILLKNQT 300

Qy        301 KLIEFLSKFQNDRTDCMSSSVPTTNSRVDLRVKPRTRGIRDLKRPAQQEA 350
              |||||||||||||:          |       ||      ||:||| |||||
Db        301 KLIEFLSKFQNDRTE-----DEQFNDEKTYLVKQ----IRNLKRAAQQEA 341
```

RESULT 2
T16651
hypothetical protein R02E12.2 - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 09-Jul-2004

```
C;Accession: T16651
R;Leimbach, D.
submitted to the EMBL Data Library, April 1996
A;Description: The sequence of C. elegans cosmid R02E12.
A;Reference number: Z18554
A;Accession: T16651
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-377 <LEI>
A;Cross-references: UNIPROT:Q21643; EMBL:U53337; NID:g1255833; PID:g1255838; PIDN:AAA961
A;Experimental source: strain Bristol N2; clone R02E12
C;Genetics:
A;Gene: CESP:R02E12.2
A;Map position: X
A;Introns: 37/3; 146/2; 225/1; 315/3
C;Superfamily: Saccharomyces hypothetical protein YKL189w

  Query Match           59.2%;  Score 1054.5;  DB 2;  Length 377;
  Best Local Similarity  62.5%;  Pred. No. 4.9e-62;
  Matches  205;  Conservative   49;  Mismatches   61;  Indels   13;  Gaps    2;

Qy          1 MPFPFGKSHKSPADIVKNLKESMAVLEK----------QDISDKKAEKATEEVSKNLVAM 50
              ||  ||||||||||:|| |:| : :|:|              |||| :|| :|||||:  :
Db          1 MPLLFGKSHKSPADVVKTLREVLTILDKLPPPKLDKDGNIQSDKKYDKALDEVSKNVAMI 60

Qy         51 KEILYGTNEKEPQTE---AVAQLAQELYNSGLLSTLVADLQLIDFEGKKDVAQIFNNILR 107
              |  :|| :  || :|    |||||||:||: :|  |:  |   :|| |||| |||||:||
Db         61 KSFIYGNDSAEPSSEHVVQVAQLAQEVYNANILPMLIKMLPKFEFECKKDVGQIFNNLLR 120

Qy        108 RQIGTRTPTVEYICTQQNILFMLLKGYESPEIALNCGIMLRECIRHEPLAKIILWSEQFY 167
              ||||||:|||||:  :  ||  |::||  |:||| ||:|||| |||: ||||||:|: ||
Db        121 RQIGTRSPTVEYLGARPEILIQLVQGGYSVPDIALTCGLMLRESIRHDHLAKIILYSDVFY 180

Qy        168 DFFRYVEMSTFDIASDAFATFKDLLTRHKLLSAEFLEQHYDRFFSEYEKLLHSENYVTKR 227
              || ||:   |||:||||:|||:| |||| : |||| :|| ||::|: ||:|:|||||:|
Db        181 TFFLYVQSEVFDISSDAFSTFKELTTRHKAIIAEFLDSNYDTFFAQYQNLLNSKNYVTRR 240

Qy        228 QSLKLLGELLLDRHNFTIMTKYISKPENLKLMMNLLRDKSRNIQFEAFHVFKVFVANPNK 287
              |||||||||||||||  |||||| |:||:||| ||||||||||:|||||||||||||||
Db        241 QSLKLLGELLLDRHNFNTMTKYISNPDNLRLMMELLRDKSRNIQYEAFHVFKVFVANPNK 300

Qy        288 TQPILDILLKNQAKLIEFLSKFQNDRTD 315
              :|| ||| :|: ||:||||:| |||||
Db        301 PKPISDILNRNREKLVEFLSEFHNDRTD 328


RESULT 3
T27129
hypothetical protein Y53C12A.4 - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 09-Jul-2004
C;Accession: T27129
R;Kershaw, J.; Lennard, N.
submitted to the EMBL Data Library, September 1997
A;Reference number: Z20315
A;Accession: T27129
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-338 <WIL>
A;Cross-references: UNIPROT:O18211; EMBL:Z99277; PIDN:CAB16486.1; GSPDB:GN00020; CESP:Y5
A;Experimental source: clone Y53C12A
C;Genetics:
A;Gene: CESP:Y53C12A.4
A;Map position: 2
A;Introns: 29/3; 103/3; 136/2; 215/1; 282/3
C;Superfamily: Saccharomyces hypothetical protein YKL189w

  Query Match           56.5%;  Score 1006;  DB 2;  Length 338;
  Best Local Similarity  60.1%;  Pred. No. 6.5e-59;
  Matches  188;  Conservative   56;  Mismatches   67;  Indels    2;  Gaps    1;

Qy          5 FGKSHKSPADIVKNLKESMAVLEKQ--DISDKKAEKATEEVSKNLVAMKEILYGTNEKEP 62
              |||: |:|||:||||:::: |:::   : |::| ||| || :| |   |  :||::  ||
Db          6 FGKADKTPADVVKNLRDALLVIDRHGTNTSERKVEKAIEETAKMLALAKTFIYGSDANEP 65

Qy         63 QTEAVAQLAQELYNSGLLSTLVADLQLIDFEGKKDVAQIFNNILRRQIGTRTPTVEYICT 122
                | | |||||:||: :|  |:  |   :|| ||||| :|||:||||||||:|||||:
Db         66 NNEQVTQLAQEVYNANVLPMLIKHLHKFEFECKKDVASVFNNLLRRQIGTRSPTVEYLAA 125

Qy        123 QQNILFMLLKGYESPEIALNCGIMLRECIRHEPLAKIILWSEQFYDFFRYVEMSTFDIAS 182
              :  ||  || ||| |:||| || |||| :||| ||:|:|:|| |  || :|:   ||||:
Db        126 RPEILITLLLGYEQPDIALTCGSMLREAVRHEHLARIVLYSEYFQRFFVFVQSDVFDIAT 185

Qy        183 DAFATFKDLLTRHKLLSAEFLEQHYDRFFSEYEKLLHSENYVTKRQSLKLLGELLLDRHN 242
              |||:|||||:|:|| : ||:|: :|||| :|  | :|||||:|||||||||||||||||
Db        186 DAFSTFKDLMTKHKNMCAEYLDNNYDRFFGQYSALTNSENYVTRRQSLKLLGELLLDRHN 245

Qy        243 FTIMTKYISKPENLKLMMNLLRDKSRNIQFEAFHVFKVFVANPNKTQPILDILLKNQAKL 302
              |: | |||: ||||| :| ||||| |||:|||||||:||||||| :|| ||| :|: ||
Db        246 FSTMNKYITSPENLKTVMELLRDKRRNIQYEAFHVFKIFVANPNKPRPITDILTRNRDKL 305

Qy        303 IEFLSKFQNDRTD 315
              :|||: | ||||:
Db        306 VEFLTAFHNDRTN 318


RESULT 4
T50117
mo25 homolog [imported] - fission yeast (Schizosaccharomyces pombe)
C;Species: Schizosaccharomyces pombe
C;Date: 09-Jun-2000 #sequence_revision 09-Jun-2000 #text_change 09-Jul-2004
C;Accession: T50117
R;Seeger, K.; Harris, D.; Wood, V.; Rajandream, M.A.; Barrell, B.G.
submitted to the EMBL Data Library, February 2000
A;Reference number: Z25039
A;Accession: T50117
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-329 <SEE>
A;Cross-references: UNIPROT:Q9P7Q8; EMBL:AL157734; PIDN:CAB75774.1; GSPDB:GN00066; SPDB
A;Experimental source: strain 972h(-); cosmid c1834
C;Genetics:
A;Gene: SPDB:SPAC1834.06c
A;Map position: 1
A;Introns: 34/3; 185/3
C;Superfamily: Saccharomyces hypothetical protein YKL189w

  Query Match           43.0%;  Score 766;  DB 2;  Length 329;
  Best Local Similarity  50.8%;  Pred. No. 3.6e-43;
  Matches  159;  Conservative   55;  Mismatches   95;  Indels    4;  Gaps    2;

Qy          1 MPFPFGKSHKSPADIVKNLKESMAVLEKQDISDKKAEKATEEVSKNLVAMKEILYGTNEK 60
              | | | |  ||  |:|: | :::  ||    :|||  |: ||||| |  ::  | || |
Db          1 MSFLFNKRPKSTQDVVRCLCDNLPKLEIN--NDKK--KSFEEVSKCLQNLRVSLCGTAEV 56

Qy         61 EPQTEAVAQLAQELYNSGLLSTLVADLQLIDFEGKKDVAQIFNNILRRQIGTRTPTVEYI 120
              ||  : |: |: ::| | |   ||  |  ::|| |||   ||: :||| : :| |||:|:
Db         57 EPDADLVSDLSFQIYQSNLPFLLVRYLPKLEFESKKDTGLIFSALLRRHVASRYPTVDYM 116

Qy        121 CTQQNILFMLLKGYESPEIALNCGIMLRECIRHEPLAKIILWSEQFYDFFRYVEMSTFDI 180
                  |  :|:   |  :|:   | :|||| ||| | :::| |  |: ||  :: |:||:
Db        117 LAHPQIFPVLVSYYRYQEVAFTAGSILRECSRHEALNEVLLNSRDFWTFFSLIQASSFDM 176

Qy        181 ASDAFATFKDLLTRHKLLSAEFLEQHYDRFFSEYEKLLHSENYVTKRQSLKLLGELLLDR 240
              |||||:||| :|  ||   |||:  |:| || :|  || |||||||||||||||:||:|
Db        177 ASDAFSTFKSILLNHKSQVAEFISYHFDEFFKQYTVLLKSENYVTKRQSLKLLGEILLNR 236

Qy        241 HNFTIMTKYISKPENLKLMMNLLRDKSRNIQFEAFHVFKVFVANPNKTQPILDILLKNQA 300
              | ::||:|||  |||||| ||||||:||||||||||:||||  |:: :::|| :|::
Db        237 ANRSVMTRYISSAENLKLMMILLRDKSKNIQFEAFHVFKLFVANPEKSEEVIEILRRNKS 296

Qy        301 KLIEFLSKFQNDR 313
              ||| :|| |  ||
Db        297 KLISYLSAFHTDR 309
```

RESULT 5
G71441
hypothetical protein - Arabidopsis thaliana
C;Species: Arabidopsis thaliana (mouse-ear cress)
A;Variety: columbia
C;Date: 03-Aug-1998 #sequence_revision 03-Aug-1998 #text_change 18-Aug-2000
C;Accession: G71441
R;Bevan, M.; Bancroft, I.; Bent, E.; Love, K.; Goodman, H.; Dean, C.; Bergkamp, R.; Dirk
P.; Wedler, H.; Wedler, E.; Wambutt, R.; Weitzenegger, T.; Pohl, T.M.; Terryn, N.; Giel
avanagh, T.; Hempel, S.; Kotter, P.; Entian, K.D.; Rieger, M.; Schaeffer, M.; Funk, B.
Nature 391, 485-488, 1998
A;Authors: Mueller-Auer, S.; Silvey, M.; James, R.; Montfort, A.; Pons, A.; Puigdomenech
erhoft, A.; Moores, T.; Jones, J.D.G.; Eneva, T.; Palme, K.; Benes, V.; Rechman, S.; Ans
C.; Chalwatzis, N.
A;Title: Analysis of 1.9 Mb of contiguous sequence from chromosome 4 of Arabidopsis thal
A;Reference number: A71400; MUID:98121113; PMID:9461215
A;Accession: G71441
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-305 <BEV>
A;Cross-references: GB:Z97343; NID:g2245073; PID:e327051; PID:g2245086
C;Genetics:
A;Map position: 4COP9-4G3845
C;Superfamily: Saccharomyces hypothetical protein YKL189w

Query Match 37.4%; Score 666.5; DB 2; Length 305;
Best Local Similarity 44.9%; Pred. No. 1.1e-36;
Matches 136; Conservative 63; Mismatches 93; Indels 11; Gaps 2;

```
Qy         42 EVSKNLVAMKEILYGTNEKEPQTEAVAQLAQELYNSGLLSTLVADLQLIDFEGKKDVAQI 101
              |:||::  :| |||| :| ||  || ||| || : :  |  |:  |  :: | :||  |:
Db          8 ELSKSIRDLKLILYGNSEAEPVAEACAQLTQEFFKADTLRRLLTSLPNLNLEARKDATQV 67

Qy        102 FNNILRRQIGTRTPTVEYICTQQNILFMLLKGYESPEIALNCGIMLRECIRHEPLAKIIL 161
              |: |:|: :|    :|: :  :::  |: |:|: ::||: | | ||||||: :|| :|
Db         68 VANLQRQQVNSRLIAADYLESNIDLMDFLVDGFENTDMALHYGTMFRECIRHQIVAKYVL 127

Qy        162 WSEQFYDFFRYVEMSTFDIASDAFATFKDLLTRHKLLSAEFLEQHYDRFFSEY-EKLLHS 220
              ||    || |:::  ||||:|| ||||:||||||   |||| :: | ||::|  ||| |
Db        128 DSEHVKKFFYYIQLPNFDIAADAAATFKELLTRHKSTVAEFLIKNEDWFFADYNSKLLES 187

Qy        221 ENYVTKRQSLKLLGELLLDRHNFTIMTKYISKPENLKLMMNLLRDKSRNIQFEAFHVFKV 280
              ||:|:||::|||||::|||| |  :||||:|  :||:::|||||: |: || |||||||:
Db        188 TNYITRRQAIKLLGDILLDRSNSAVMTKYVSSMDNLRILMNLLRESSKTIQIEAFHVFKL 247

Qy        281 FVANPNKTQPILDILLKNQAKLIEFLSKFQNDRTDCMSSSVPTTNSRVDLRVKPRTRGIR 340
              |||| ||   | :||: |: ||:  |:  : |: |           | |       | |
Db        248 FVANQNKPSDIANILVANRNKLLRLLADIKPDKED----------ERFDADKAQVVREIA 297

Qy        341 DLK 343
              :||
Db        298 NLK 300
```

RESULT 6
B84448
hypothetical protein At2g03410 [imported] - Arabidopsis thaliana
C;Species: Arabidopsis thaliana (mouse-ear cress)
C;Date: 02-Feb-2001 #sequence_revision 02-Feb-2001 #text_change 09-Jul-2004
C;Accession: B84448
R;Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.;
M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; VanAken, S.E.; Umayam, L.; Tallon, L.
euss, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter, J
Nature 402, 761-768, 1999
A;Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.
A;Reference number: A84420; MUID:20083487; PMID:10617197
A;Accession: B84448
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-348 <STO>
A;Cross-references: UNIPROT:Q9ZQ77; GB:AE002093; NID:g4335758; PIDN:AAD17435.1; GSPDB:GN
C;Genetics:
A;Gene: At2g03410
A;Map position: 2
C;Superfamily: Saccharomyces hypothetical protein YKL189w

Query Match 34.5%; Score 614; DB 2; Length 348;
Best Local Similarity 39.4%; Pred. No. 3.7e-33;
Matches 125; Conservative 74; Mismatches 112; Indels 6; Gaps 3;

```
Qy          5 FGKSHKSPADIVKNLKESMAVLEKQ----DISDKKAEKATEEVSKNLVAMKEILYGTNEK 60
              |    : | :||:  :: :|: | :    |  : |       |: :|:  :| |||| |
Db          5 FKNKSRLPGEIVRQTRDLIALAESEEEETDARNSKRLGICAELCRNIRDLKSILYGNGEA 64

Qy         61 EPQTEAVAQLAQELYNSGLLSTLVADLQLIDFEGKKDVAQIFNNILRRQIGTRTPTVEYI 120
              ||  ||   | || : :  |  |:  :  :| | :||  ||  |: ::|:  |    ||:
Db         65 EPVPEACLLLTQEFFRADTLRPLIKSIPKLDLEARKDATQIVANLQKQQVEFRLVASEYL 124

Qy        121 CTQQNILFMLLKGYE-SPEIALNCGIMLRECIRHEPLAKIILWSEQFYDFFRYVEMSTFD 179
               :  ::: |::| :   |:||:   ||:||:||: :|| || |:    || ||:: ||
Db        125 ESNLDVIDSLVEGIDHDHELALHYTGMLKECVRHQVVAKYILESKNLEKFFDYVQLPYFD 184

Qy        180 IASDAFATFKDLLTRHKLLSAEFLEQHYDRFFSEYE-KLLHSENYVTKRQSLKLLGELLL 238
              :|:||   |::||||||   ||:| ::|: ||:|| |||   :| ||||: ||||::|:
Db        185 VATDASKIFRELLTRHKSTVAEYLAKNYEWFFAEYNTKLLEKGSYFTKRQASKLLGDVLM 244

Qy        239 DRHNFTIMTKYISKPENLKLMMNLLRDKSRNIQFEAFHVFKVFVANPNKTQPILDILLKN 298
              || |  :| ||:|  :||::|||||: ::||| ||||:||:|||| || : |: ||: |
Db        245 DRSNSGVMVKYVSSLDNLRIMMNLLREPTKNIQLEAFHIFKLFVANENKPEDIVAILVAN 304

Qy        299 QAKLIEFLSKFQNDRTD 315
              : |::   :  : :: |
Db        305 RTKILRLFADLKPEKED 321
```

RESULT 7
S34681
hypothetical protein YKL189w - yeast (Saccharomyces cerevisiae)
C;Species: Saccharomyces cerevisiae
C;Date: 30-Sep-1993 #sequence_revision 30-Sep-1993 #text_change 09-Jul-2004
C;Accession: S34681; S33963; S38021; S38026
R;Wieman, S.; Voss, H.; Schwagaer, C.; Rupp, T.; Stegemann, J.; Zimmermann, J.; Grothues,
submitted to the EMBL Data Library, July 1993
A;Description: Sequencing and analysis of 51.5 kilobases on the left arm of chromosome XI
A;Reference number: S34679
A;Accession: S34681
A;Molecule type: DNA
A;Residues: 1-399 <WIE>
A;Cross-references: UNIPROT:P32464; EMBL:X74151; NID:g450365; PIDN:CAA52249.1; PID:g3952
A;Experimental source: strain S288C
R;Cheret, G.; Mattheakis, L.C.; Sor, F.
Yeast 9, 661-667, 1993
A;Title: DNA sequence analysis of the YCN2 region of chromosome XI in Saccharomyces cere
A;Reference number: S33960; MUID:93348778; PMID:8394042
A;Accession: S33963
A;Molecule type: DNA
A;Residues: 1-399 <CHE>
A;Cross-references: GB:X69765; NID:g296985; PIDN:CAA49422.1; PID:g296989
R;Wiemann, S.; Voss, H.; Schwager, C.; Rupp, T.; Grothues, D.; Sensen, C.; Stegemann, J.;
submitted to the Protein Sequence Database, March 1994
A;Reference number: S37825
A;Accession: S38021
A;Molecule type: DNA
A;Residues: 1-399 <WI2>
A;Cross-references: EMBL:Z28189; NID:g486334; PIDN:CAA82032.1; PID:g486335; MIPS:YKL189w
A;Experimental source: strain S288C
R;Maia e Silva, A.; Bossier, P.; Vilela, C.; Fernandes, L.; Soares, H.; Guerreiro, P.; Ro
submitted to the Protein Sequence Database, March 1994
A;Reference number: S38024
A;Accession: S38026
A;Molecule type: DNA
A;Residues: 1-399 <MAI>
A;Cross-references: EMBL:Z28189; NID:g486334; PIDN:CAA82032.1; PID:g486335; MIPS:YKL189w
A;Experimental source: strain S288C
C;Genetics:
A;Gene: SGD:HYM1

```
A;Cross-references: SGD:S0001672
A;Map position: 11L
C;Superfamily: Saccharomyces hypothetical protein YKL189w

  Query Match             28.1%;  Score 501;  DB 2;  Length 399;
  Best Local Similarity   33.8%;  Pred. No. 1.2e-25;
  Matches  110;  Conservative   68;  Mismatches  137;  Indels   10;  Gaps    3;

Qy          1 MPFPPFGKSHKSPADIVKNLKESMAVLEKQDISDKKAEKATEEVSKNLVAMKEILYGTNEK 60
              | | : |: |:|:|  : : | :        ::      |  || :| |:  |  : | :
Db         12 MAFWWKKNPKTPSDYARLIIEQLNKFSSPSLTQDNKRKVQEECTKYLIGTKHFIVGDTDP 71

Qy         61 EPQTEAVAQLAQELYNSGLLSTLVADLQLIDFEGKKDVAQIFNNILRRQIGTRTPTVEYI 120
               |  ||: :|   :: : :   |:     ::|| :::   ||:  |      :  ||:|:
Db         72 HPTPEAIDELYTAMHRADVFYELLLHFVDLEFEARRECMLIFSICLGYSKDNKFVTVDYL 131

Qy        121 CTQQNILFMLLKGYE-------SPEIALNCGIMLRECIRHEPLAKIILWSEQFYDFFRYV 173
               :|   : ::|: |          :| |  | |: |||::| | :|||   | : || :
Db        132 VSQPKTISLMLRTAEVALQQKGCQDIFLTVGNMIIECIKYEQLCRIILKDPQLWKFFEFA 191

Qy        174 EMSTFDIASDAFATFKDLLTRH-KLLSAEFLEQHYD--RFFSEYEKLLHSENYVTKRQSL 230
              ::  |:|::::         | | ||:| ||     :  ||      ||:   :|||||||
Db        192 KLGNFEISTESLQILSAAFTAHPKLVSKEFFSNEINIIRFIKCINKLMAHGSYVTKRQST 251

Qy        231 KLLGELLLDRHNFTIMTKYISKPENLKLMMNLLRDKSRNIQFEAFHVFKVFVANPNKTQP 290
              |||  |:: | |  :|  ||: |||||:| |: |||:|:| |||:|||| |||| |::|
Db        252 KLLASLIVIRSNNALMNIYINSPENLKLIMTLMTDKSKNLQLEAFNVFKVMVANPRKSKP 311

Qy        291 ILDILLKNQAKLIEFLSKFQNDRTD 315
              : |||:||: ||: :   |  |  |
Db        312 VFDILVKNRDKLLTYFKTFGLDSQD 336


RESULT 8
T33477
hypothetical protein T27C10.3 - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 09-Jul-2004
C;Accession: T33477
R;Zhu, H.J.; Graves, T.; Hawkins, M.
submitted to the EMBL Data Library, October 1998
A;Description: The sequence of C. elegans cosmid T27C10.
A;Reference number: Z21354
A;Accession: T33477
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-339 <ZHU>
A;Cross-references: UNIPROT:Q9TZM2; EMBL:AF098504; PIDN:AAC67411.1; GSPDB:GN00019; CESP:
A;Experimental source: strain Bristol N2; clone T27C10
C;Genetics:
A;Gene: CESP:T27C10.3
A;Map position: 1
A;Introns: 72/3; 120/3; 233/3; 295/1

  Query Match              7.7%;  Score 137.5;  DB 2;  Length 339;
  Best Local Similarity   20.0%;  Pred. No. 0.071;
  Matches   36;  Conservative   43;  Mismatches   82;  Indels   19;  Gaps    3;

Qy        178 FDIASDAFATFKDLLTRHKLLSAEFLEQHYDRFFSEYEKLLHSENYVTKRQSLKLLGELL 237
              ||:    | | : :   :   :  |:: :  ||      ||:   |:  : :| | | ||
Db        109 FDVIQGTFDTLQIIFFTNHESANNFIKNNLPRFMQTLHKLIACSNFFIQAKSFKFLNELF 168

Qy        238 LDRHNFTIMTKYISKPENLKLMMNLLRDKSRNIQFEAFHVFKVFVANPNKTQPILDILLK 297
                : |:   : :::|  :||::  ::     ::  |  : ::|: ||   :  : : : :
Db        169 TAQTNYETRSLWMAEPAFIKLVVLAIQSNKHAVRSRAVSILEIFIRNPRNSPEVHEFIGR 228

Qy        298 NQAKLIEFLSKFQNDRTDCMSSSVP------TTNSRVDLRVKPRTRGIR--DLKRPAQQE 349
              |:  || |           :| |     : | : |  :     :   |::||  ||
Db        229 NRNVLIAF-----------FFNSAPIHYYQGSPNEKEDAQYARMAYKLLNWDMQRPFTQE 277


RESULT 9
T03057
hypothetical protein 032R - Chilo iridescent virus
C;Species: Chilo iridescent virus
C;Date: 24-Mar-1999 #sequence_revision 24-Mar-1999 #text_change 09-Jul-2004
C;Accession: T03057
R;Bahr, U.; Tidona, C.A.; Darai, G.
Virus Genes 15, 235-245, 1997
A;Title: The DNA sequence of Chilo iridescent virus between the genome coordinates 0.101
A;Reference number: Z14834; MUID:98141693; PMID:9482589
A;Accession: T03057
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-602 <BAH>
A;Cross-references: UNIPROT:O55720; EMBL:AF003534; NID:g2738385; PIDN:AAB94431.1; PID:g2
C;Superfamily: Chilo iridescent virus hypothetical protein 032R

  Query Match              6.9%;  Score 122;  DB 2;  Length 602;
  Best Local Similarity   22.1%;  Pred. No. 1.5;
  Matches   75;  Conservative   57;  Mismatches  145;  Indels   62;  Gaps   10;

Qy          4 PFGKSHKSPADIVKNLK------------ESMAVLEKQDISDKKAEKAT-------EEV 43
              ||  :   | |::||::             |  :  ||  :: ||::  |       | :
Db        107 PFLPNFMKPYDLIKNVRADPRQKNPFVTLIEEKSGSEKMRLTTKKSKPKTFSDMALFEYI 166

Qy         44 SKNLVAMKEILYGTNEKEPQTEAVAQLAQELYNSG--LLSTLVADLQLIDFEGKKDVAQI 101
              : |:   : |        |       : :  |::  |   :::|:  | :     :| :: |
Db        167 NSNMTLAELISINVKGKRQTRVTLNEEIQKIKASDFKIINSLLNQLMIAILIAQKKLSFI 226

Qy        102 -----FNNILRRQIGTRTPTVEYICTQQNILFMLLKGYESPEIALNCGIMLRECIRHEPL 156
                  |:|:|  :   ||  : |:   :| : ||  |    | :: |    | :  ||
Db        227 HNDLHFDNVLICKCLQRT-FMLYVFEYNSICYALLPTYGYYPIIIDYGFSFSEDLIGGPL 285

Qy        157 AKIILWSEQFYDFFRYVEMSTFDIASDAFATFKDLLTRHKLLSAEFLEQHYDRFFSEYEK 216
                |  : : |   :|          | |  || :|||      :|     | |
Db        286 LTGIHHNNKGYMNHQY----------DEFTDFKTMLTRLSYSGYQFGLDKKDAF------ 329

Qy        217 LLHSENYVTKRQSL---KLLGELLLDRHNFTIMTKYISKPENLKLMMNLLRD---KSRNI 270
                          |||   ||: :| :|:      || ||  : |   : :  |    || |
Db        330 ----------QSLIFDKLISKLPIDKQTGWDETKDISVSKQLVRHIRIFVDDYLKSINR 378

Qy        271 QFEAFHVFKVFVANPNKTQPILDILLKNQAKLIEFLSKF 309
              :  |  :   : :   :  || :  ||    |:| |: |
Db        379 E-SFFQKYDYEMVDMIGSLIILPLRKKNTENLVETLAIF 416


RESULT 10
D97848
hypothetical protein RC1188 [imported] - Rickettsia conorii (strain Malish 7)
C;Species: Rickettsia conorii
C;Date: 30-Sep-2001 #sequence_revision 30-Sep-2001 #text_change 09-Jul-2004
C;Accession: D97848
R;Ogata, H.; Audic, S.; Renesto-Audiffren, P.; Fournier, P.E.; Barbe, V.; Samson, D.; Ro
Science 293, 2093-2098, 2001
A;Title: Mechanisms of Evolution in Rickettsia conorii and Rickettsia prowazekii.
A;Reference number: A97700; MUID:21442074; PMID:11557893
A;Accession: D97848
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-365 <KUR>
A;Cross-references: UNIPROT:Q92GD5; GB:AE006914; PIDN:AAL03726.1; PID:g15620317; GSPDB:G
C;Genetics:
A;Gene: RC1188

  Query Match              6.6%;  Score 117;  DB 2;  Length 365;
  Best Local Similarity   20.1%;  Pred. No. 1.7;
  Matches   74;  Conservative   60;  Mismatches  127;  Indels  108;  Gaps   14;

Qy         16 VKNLKESMAVLEKQDISDKKAEKATEEVSKNLVAMKEILYGTNEKEPQTEAVAQLAQELY 75
              : ||||:      |     ::  |||  :  :  | : :  |      |||      :||    |
Db         60 ISNLKETAETKLKHYEVEQSIEKARVQEEEFLKSQRAIKLAEEEKE-----IAQRIAALN 114

Qy         76 N--SGLLSTLVADLQLIDFEGKKDVAQIFNNILRRQIGTRTPTVEYICTQQNILFMLLKG 133
              |  :  :  :  | : |: |  | : :| | : :  |
Db        115 NLHNEFIKNITKDTKRIB-ESNKRLDKIINKLEKENI---------------------- 150
```

```
Qy        134 YESPEIALNCGIMLRECIRHEPLAKIILWSEQFYDFFRYVEMSTFDIASDAFATFKDLLT 193
Db        151 -----------------IDHEELNREILTHEEIFKLNQ--------------AYKVLHK 178

Qy        194 RHKLLSAEFLEQHYDRFFSEYEKLLHS------ENYVTKRQSLKLLGELLLDRHNFTIMT 247
Db        179 NHKKITEEHKQAHTE--LNELNKTITNLSQQLQKKGLTPKKVKELKGELEFHQEMLKIHK 236

Qy        248 KYISKPENLKLMMN---LLRDKSRNIQFEAFHVFKVFVANPNKTQP--ILD-----ILLK 297
Db        237 AHVEKIENSKKIVDQEIIKYEKEHNINKDKIKKLGSDIKERYKKEPEKYLDAYEKYLVLK 296

Qy        298 NQAKLI------------EFLSKFQNDRTDCMS------SSVPTTNSR-------VDLR 331
Db        297 HQHKAIKTDGVVKDIEHHNKVLNKINVDKTKSLANKIREQTQVKDTNTRSLSPARTPNIK 356

Qy        332 V-KPRTRGI 339
Db        357 VTNNKTRGI 365
```

RESULT 11
A90551
conserved hypothetical protein MYPU_3130 [imported] - Mycoplasma pulmonis (strain UAB CT
C;Species: Mycoplasma pulmonis
C;Date: 24-May-2001 #sequence_revision 24-May-2001 #text_change 09-Jul-2004
C;Accession: A90551
R;Chambaud, I.; Heilig, R.; Ferris, S.; Barbe, V.; Samson, D.; Galisson, F.; Moszer, I.;
Nucleic Acids Res. 29, 2145-2153, 2001
A;Title: The complete genome sequence of the murine respiratory pathogen Mycoplasma pulm
A;Reference number: A99512; MUID:21267165; PMID:11353084
A;Accession: A90551
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-2819 <KUR>
A;Cross-references: UNIPROT:Q98QP8; GB:AL445566; PID:g14089727; PIDN:CAC13486.1; GSPDB:G
A;Experimental source: strain UAB CTIP
C;Genetics:
A;Gene: MYPU_3130
A;Genetic code: SGC3

  Query Match 6.5%; Score 115; DB 2; Length 2819;
  Best Local Similarity 19.0%; Pred. No. 31;
  Matches 83; Conservative 87; Mismatches 127; Indels 140; Gaps 24;

```
Qy          5 FGKSHKSPADIVKNL----------KESMAVLEK-QDISDKKAEKATE--EVSKNLVAM 50
Db       1404 FGKT-----NIISNLYDLVLSALSINQESSSVLTKVENITHKALDQQNSYFAVIKTLLSV 1458

Qy         51 KEILYGTNEKEPQTEAVAQLAQELY--NSGLLSTLVADL--QLIDFEG--KKDVAQIFNN 104
Db       1459 KE---GDVDKSDEFKAIIKKFLNLFAKNDYLLTKFISPIVDKTIDFDSATKSSIVKLIKN 1515

Qy        105 ILRRQ------------IGTRTPTVEY----------ICTQQNILFM---------- 129
Db       1516 VLSDEQNLKFSAQIIDTIFANKNKYLAYENLNDFLIDALVDNKENIINLSSNTLGKIKTD 1575

Qy        130 -----LLKGYESPEIALNCGIMLRECIR------HEPLAKIILWSEQFY-----DFFRYV 173
Db       1576 DEFKNIIKSFITKNLMTSDSSITEEEINTIVKSAHELIS--IVDSSNFLNQLLKDFFALL 1633

Qy        174 EMSTFDIASDA--FATFKDLLTRHKLLSAEFLEQHYDRFFSEYEKLLHSENYVTKRQSLK 231
Db       1634 EKNKKDSKSTSLEFSTFM------KKSISEFLDAE-NNYLALFKKVLSSNLLSTQS---- 1682

Qy        232 LLGELLLDRHNFT-----IMTKYISKPENLKLMMNLLRDKSRNIQF------EAFHVFKV 280
Db       1683 -------DQSSFTKSISSILRVFLRKEKALELM---LKKQFRNFELHDIKVDDAINLIK- 1731

Qy        281 FVANPNKTQPILDILLKN-------------QAKLIEFLSKFQNDRTDCMSSSV----- 321
Db       1732 FIFRDNQVIDFIETIINRVVSQKEEYSKLNSYQEILYKFLSSNKNETLNFFKKQIGEIKN 1791

Qy        322 -----PTTNSRVDLRVK 333
Db       1792 SKLINPLAKSFIESELK 1808
```

RESULT 12
T18372
repeat organellar protein - Plasmodium chabaudi
C;Species: Plasmodium chabaudi
C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 09-Jul-2004
C;Accession: T18372
R;Werner, E.B.; Taylor, W.R.; Holder, A.A.
Mol. Biochem. Parasitol. 94, 185-196, 1998
A;Title: A Plasmodium chabaudi protein contains a repetitive region with a predicted spe
A;Reference number: Z18922; MUID:98418765; PMID:9747969
A;Accession: T18372
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-1939 <WER>
A;Cross-references: UNIPROT:Q25662; EMBL:U43145; NID:g1151157; PID:g1151158; PIDN:AAC634(

  Query Match 6.4%; Score 113.5; DB 2; Length 1939;
  Best Local Similarity 19.4%; Pred. No. 24;
  Matches 57; Conservative 57; Mismatches 113; Indels 67; Gaps 11;

```
Qy         18 NLKESMAVLEKQDISDKKAEKATEEVS--KNLVAMKEILYG--TNEKEPQTEAVAQLAQE 73
Db       1433 NLNKNVEDKTGDDINCEKNNDQAKEISYLKDEIKKISMLYGEELNRKNSYDEKVKNLTNE 1492

Qy         74 LYNSGLLSTLVADLQLIDFEGKKDVAQI-------------------FNNILRRQIGTR 113
Db       1493 L----------KELKIRNKKGEEAIAELNKLKNIKEKNKSVKQNDESSSNNIITKD-GDK 1541

Qy        114 TPTVEYICTQQNI-----LFMLLKGYESPEIALNCGIMLRECIRHEPLAK--IILWSEQF 166
Db       1542 TP--EYVSNDDKIQKDWKANLVLKLKEKPDLWDNINSLEKENFRVMSIVKENKNVQNDKI 1599

Qy        167 YDFFRYVEMSTFDIASDAFATFKDLLTRHKLLSAEFLEQHYDRFFSEYEKLLHSENYVTK 226
Db       1600 VGIYSYFKKCEKELKNDMLVI---CLVLKDILSILFLNDNFVNLFEKIDKILWKQMYIPT 1656

Qy        227 ----------------RQSLKLLGELLLD----RHNFTIMTKYISKPENLKLMM 260
Db       1657 EIRILFLRYFSFLDKLRNYVKCVNEEYVNNERYEYSWALFQTYLETASNLKKEM 1710
```

RESULT 13
H64574
DNA topoisomerase I - Helicobacter pylori (strain 26695)
C;Species: Helicobacter pylori
C;Date: 09-Aug-1997 #sequence_revision 09-Aug-1997 #text_change 09-Jul-2004
C;Accession: H64574
R;Tomb, J.F.; White, O.; Kerlavage, A.R.; Clayton, R.A.; Sutton, G.G.; Fleischmann, R.D.;
 Peterson, S.; Loftus, B.; Richardson, D.; Dodson, R.; Khalak, H.G.; Glodek, A.; McKenney
son, J.D.; Kelley, J.M.; Cotton, M.D.; Weidman, J.M.; Fujii, C.; Bowman, C.; Watthey, L.
Nature 388, 539-547, 1997
A;Authors: Wallin, E.; Hayes, W.S.; Borodovsky, M.; Karpk, P.D.; Smith, H.O.; Fraser, C.M
A;Title: The complete genome sequence of the gastric pathogen Helicobacter pylori.
A;Reference number: A64520; MUID:97394467; PMID:9252185
A;Accession: H64574
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-677 <TOM>
A;Cross-references: UNIPROT:O25188; GB:AE000559; GB:AE000511; NID:g2313536; PIDN:AAD07502
C;Superfamily: DNA topoisomerase I

  Query Match 6.3%; Score 112.5; DB 2; Length 677;
  Best Local Similarity 19.4%; Pred. No. 7.4;
  Matches 84; Conservative 80; Mismatches 141; Indels 129; Gaps 20;

```
Qy          3 FPFGKSHKSPADIVKNLKE---SMAVLE--KQDISDKK------AEKATEEVSKNLVAMK 51
```

```
Db        222 FKF-KDKNEASQFLKDLKDGLGSMSVLVSLKESLSNKEPKKPFTTSKLLSQASKSL---- 276

Qy         52 EILYGTNEKEPQTEAVAQLAQELYNSGLLSTLVADLQLIDFEGKKDVAQIFNNILRRQIG 111

Db        277 --------KIP-TKEIAQLAQKLFEAGLITYHRTDSEFLSPEYLKEHEVFFEPIY----- 322

Qy        112 TRTPTV----EYIC-------------TQQNILFMLLKGYESPEIALNCGIMLRECIRH 153

Db        323 ---PSVYQYREYKAGKNSQAEAHEAIRITHPHALKDLEKVCSDAKISEELALKLYQLIYT 379

Qy        154 EPL---AKIILWSEQFYDFFRYVEMSTFDIA-----SDAFATFKDLLTRHKLLSAEFLEQ 205

Db        380 NTICSQSRNALYNQ--YDCIFKIKSESFKLSFKLLKEKGFLEIEELIQGKEEINREEQES 437

Qy        206 HYDRFF-------------------------SEYEKLLHSENYVTKRQSLKLLGELLL 238

Db        438 EIENFSLKENDSVPLKEVFIKKIEKPSPKPYKESAFIPLLESEG-IGRPSTYASFLDLLL 496

Qy        239 DRHNFTIMTK---------------YISKPENLKLMMNLLRDKSRNIQFEAFHVFKVFVA 283

Db        497 KRKYISIDTKTNAITPTSQGLEVISFFKKDKEVDFIALTSKDKSK-------------LG 543

Qy        284 NPNKT-QPILDILLKNQAK----LIEFLSKFQND----RTDCMSSSVPTTNSRVDLRVKP 334

Db        544 NTTKQFEECLDLIMRGEASYEKFMLEVISKLKSTAKFYQTQSTDNNMPTDK---QLELID 600

Qy        335 RTRGIRDLKRPAQQ 348

Db        601 KICKDKKLQKPSQE 614


RESULT 14
T18489
hypothetical protein C0820w - malaria parasite (Plasmodium falciparum)
C;Species: Plasmodium falciparum
C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 09-Jul-2004
C;Accession: T18489
R;Lawson, D.; Bowman, S.; Barrell, B.
submitted to the EMBL Data Library, August 1997
A;Reference number: Z18935
A;Accession: T18489
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-4981 <LAW>
A;Cross-references: UNIPROT:O77372; EMBL:Z98551; NID:e1331903; PID:e1331910; PIDN:CAB111
C;Genetics:
A;Map position: 3
A;Note: C0820w

  Query Match              6.2%;  Score 111;  DB 2;  Length 4981;
  Best Local Similarity   20.1%;  Pred. No. 1.2e+02;
  Matches   59;  Conservative   49;  Mismatches  104;  Indels   82;  Gaps   10;

Qy        100 QIFNNILRRQIGTRTPTVEYICTQQNILFMLLKGYES----PEIALNCGIMLRECIRHEP 155

Db       1521 KIFLNYIRKEKEHIYNFQTYITTFFFTLLEMLKNIQSDKDYEKTLYNIGIINMMCVAVKF 1580

Qy        156 LA-----KII-------------LWSEQ-----FYDFFRYVEMSTFDIASDAFATFKDLL 192

Db       1581 LKEKCPNKIISIFFKLLEEVPIPLGEEEDKVVTLYKNYSYIYRTHF------FNLYKDKC 1634

Qy        193 TRHKLLSAEFLEQHYDRFFSEYEKLLHSENYVTKRQSLKLLGELLLDRHNFTIMTKYIS- 251

Db       1635 T-----------TLYTKNCMEYIKMMRNNIWLNKKKN-----DSQVDFNNFIILSNMYNN 1678

Qy        252 ----KPENLKLMMNLLRDKSRNIQFEAFHVFKVFVAN------------------PNKT 288

Db       1679 SDTFKKKNIQNVLQYLNDYQQVYKKNVVHIFNMYISNVVTHKEMENGMYRKEGNKRENNT 1738

Qy        289 QPILDILLKNQ---------AKLIEFLSKFQNDRTDCMSSSVPTTNSRVDLRVK 333

Db       1739 YDKINIDDKNNFGQPIYKYVDKLTSKRNTLDNSSINCLESSYVTKGSNTKIRIK 1792


RESULT 15
G70163
hypothetical protein BB0512 - Lyme disease spirochete
C;Species: Borrelia burgdorferi (Lyme disease spirochete)
C;Date: 13-Feb-1998 #sequence_revision 13-Feb-1998 #text_change 09-Jul-2004
C;Accession: G70163
R;Fraser, C.M.; Casjens, S.; Huang, W.M.; Sutton, G.G.; Clayton, R.; Lathigra, R.; White
son, D.; Peterson, J.; Kerlavage, A.R.; Quackenbush, J.; Salzberg, S.; Hanson, M.; Vugt,
; Bowman, C.; Garland, S.; Fujii, C.; Cotton, M.D.; Horst, K.; Roberts, K.; Hatch, B.
Nature 390, 580-586, 1997
A;Authors: Smith, H.O.; Venter, J.C.
A;Title: Genomic sequence of a Lyme disease spirochaete, Borrelia burgdorferi.
A;Reference number: A70100; MUID:98065943; PMID:9403685
A;Accession: G70163
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-2166 <KLE>
A;Cross-references: UNIPROT:O51465; GB:AE001153; GB:AE000783; NID:g2688419; PIDN:AAC6687
A;Experimental source: strain B31

  Query Match              6.1%;  Score 109;  DB 2;  Length 2166;
  Best Local Similarity   22.5%;  Pred. No. 55;
  Matches   80;  Conservative   68;  Mismatches  114;  Indels   94;  Gaps   23;

Qy         17 KNLKES--MAVLEKQDISDKKAEKATEEVSK---NLVAMKEILYGTNEKEPQTEAVAQLA 71

Db       1623 KNIGESLNLKVLDLEKFVDFKLEKIDEKVNKKTEDILIQAEVKFLTQQKDLE-DKIFELN 1681

Qy         72 QELYNSGLLSTLVADL-----QLID--------FEGK-----KDVAQIFNNI--LRRQIG 111

Db       1682 QKLEHE--FTTLSSNLDKVRREMVDVISSDKESFEGQIELINKNISEFSEKISLYRNNIE 1739

Qy        112 TRTPTVEYICTQQNILFMLLKGYESPEIALNCGIM---LRECIRH-------------EP 155

Db       1740 TSIEN-EY-------------NSFSKSIKSDLGLLEDELKKSLKHSTSEIETIKSGLQEQ 1785

Qy        156 LAKI-ILWSEQFYDFFRYVEMSTFDIASDA----------FATFKDLLTRHKL-LSAEFL 203

Db       1786 IDKFEVEFKKNHKELLKEVDNNILELESKILNCDVQFNKFISEIKDNLVEYKSDLRAEF- 1844

Qy        204 EQHYDRFFSEYEKLLHSENYVTKRQSLKLLGELLLDRHNFTIMTKYISKP------ENLK 257

Db       1845 EDSYDKINFQIENQI--ENF------KKLDSE--LEKNNSIFLEAYSLKDKLEKLWETLK 1894

Qy        258 LMMNLLRDKSRNIQFEAFHVFKVFVANPNKTQPILDILLKNQAKL-IEFLSKFQND 312

Db       1895 NEIGLAQEYKNNFE----NVNKEFYNIQKETLGIIEIF--NELKLEQESIKSIKND 1944


Search completed: October 22, 2004, 09:13:07
Job time : 41 secs
```

GenCore version 5.1.6

OM protein - protein search, using sw model

Run on: October 22, 2004, 08:56:03 ; Search time 191 Seconds
(without alignments)
1054.351 Million cell updates/sec

Title: US-10-089-688-2
Perfect score: 1781
Sequence: 1 MPFPPFGKSHKSPADIVKNLK.....RVKPRTRGIRDLKRPAQQEA 350

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1825181 seqs, 575374646 residues

Total number of hits satisfying chosen parameters: 1825181

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : UniProt_02:*
1: uniprot_sprot:*
2: uniprot_trembl:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	1635.5	91.8	341	1	CB39_HUMAN	Q9y376 homo sapien
2	1619.5	90.9	341	2	Q8VDZ8	Q8vdz8 mus musculu
3	1614.5	90.7	341	1	CB39_MOUSE	Q06138 mus musculu
4	1614.5	90.7	341	2	Q6IP72	Q6ip72 xenopus lae
5	1614.5	90.7	341	2	AAH72045	Aah72045 xenopus l
6	1590.5	89.3	341	2	Q7Z4X0	Q7z4x0 homo sapien
7	1537.5	86.3	341	2	Q6IQL2	Q6iql2 brachydanio
8	1537.5	86.3	341	2	AAH71393	Aah71393 brachydan
9	1339.5	75.2	337	2	Q6PAB4	Q6pab4 xenopus lae
10	1339.5	75.2	337	2	AAH60384	Aah60384 xenopus l
11	1298.5	72.9	337	1	C39L_MOUSE	Q9db16 mus musculu
12	1297.5	72.9	337	1	C39L_HUMAN	Q9h9s4 homo sapien
13	1287.5	72.3	337	2	AAQ93064	Aaq93064 homo sapi
14	1287	72.3	343	2	Q803V8	Q803v8 brachydanio
15	1163.5	65.3	332	2	Q7PF07	Q7pf07 anopheles g
16	1163.5	65.3	335	2	Q7PRN5	Q7prn5 anopheles g
17	1123	63.1	339	1	MO25_DROME	P91891 drosophila
18	1123	63.1	339	2	AAM75031	Aam75031 drosophil
19	1054.5	59.2	377	2	Q7Z2A5	Q7z2a5 caenorhabdi
20	1054.5	59.2	636	2	Q21643	Q21643 caenorhabdi
21	1006	56.5	338	1	MO2M_CAEEL	O18211 caenorhabdi
22	944.5	53.0	205	2	Q8K312	Q8k312 mus musculu
23	766	43.0	329	1	YFV6_SCHPO	Q9p7q8 schizosacch
24	764.5	42.9	321	1	DE76_CHLPR	Q9xfy6 chlorella p
25	726	40.8	390	2	Q6CAN7	Q6can7 yarrowia li
26	711.5	39.9	343	1	MO2N_ARATH	Q9fgk3 arabidopsis
27	711.5	39.9	343	2	AAL16128	Aal16128 arabidops
28	705.5	39.6	343	1	MO2M_ARATH	Q9m0m4 arabidopsis
29	705.5	39.6	343	2	AAL06959	Aal06959 arabidops
30	686	38.5	338	2	Q6BSQ5	Q6bsq5 debaryomyce
31	669.5	37.6	345	2	Q8L9L9	Q8l9l9 arabidopsis
32	651.5	36.6	337	2	Q7XIQ4	Q7xiq4 oryza sativ
33	623	35.0	384	1	HYMA_EMENI	O60032 emericella
34	614	34.5	348	1	MO2L_ARATH	Q9zq77 arabidopsis
35	614	34.5	348	2	AAR24657	Aar24657 arabidops
36	584	32.8	370	2	Q873K5	Q873k5 neurospora
37	571	32.1	364	2	Q6CWP5	Q6cwp5 kluyveromyc
38	560	31.4	362	2	Q757C0	Q757c0 ashbya goss
39	560	31.4	362	2	AAS52777	Aas52777 ashbya go
40	511	28.7	422	2	Q6FWG7	Q6fwg7 candida gla
41	501	28.1	399	1	HYM1_YEAST	P32464 saccharomyc
42	455.5	25.6	332	2	Q7QTG6	Q7qtg6 giardia lam
43	327	18.4	103	2	Q8K038	Q8k038 mus musculu
44	283.5	15.9	509	2	Q7YYL6	Q7yyl6 cryptospori
45	137.5	7.7	339	1	MO2L_CAEEL	Q9tzm2 caenorhabdi

ALIGNMENTS

RESULT 1
CB39_HUMAN
ID CB39_HUMAN STANDARD; PRT; 341 AA.
AC Q9Y376;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 01-OCT-2004 (Rel. 45, Last annotation update)
DE Calcium binding protein 39 (Mo25 protein) (CGI-66).
GN Name=CAB39; Synonyms=MO25;
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20272150; PubMed=10810093;
RA Lai C.-H., Chou C.-Y., Ch'ang L.-Y., Liu C.-S., Lin W.-C.;
RT "Identification of novel human genes evolutionarily conserved in
RT Caenorhabditis elegans by comparative proteomics.";
RL Genome Res. 10:703-713(2000).
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=Hypothalamus;
RA Jin W., Shi J., Ren S., Gu J., Fu S., Huang Q., Dong H., Yu Y., Fu G.,
RA Wang Y., Chen Z., Han Z.;
RT "A novel gene expressed in the human hypothalamus.";
RL Submitted (DEC-1998) to the EMBL/GenBank/DDBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC TISSUE=Duodenum;
RX MEDLINE=22388257; PubMed=12477932; DOI=10.1073/pnas.242603899;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahey J., Helton E., Ketteman M., Madan A., Rodrigues S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smailus D.E.,
RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length human
RT and mouse cDNA sequences.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN [4]
RP X-RAY CRYSTALLOGRAPHY (1.85 ANGSTROMS) OF 10-339.
RX PubMed=14730349; DOI=10.1038/nsmb716;

```
RA   Milburn C.C., Boudeau J., Deak M., Alessi D.R., van Aalten D.M.;
RT   "Crystal structure of MO25 alpha in complex with the C terminus of the
RT   pseudo kinase STE20-related adaptor.";
RL   Nat. Struct. Mol. Biol. 11:193-200(2004).
CC   -!- FUNCTION: Together with the STE20-related adaptor-alpha (STRAD
CC       alpha) pseudo kinase, forms a regulatory complex capable of
CC       stimulating the activity of STK11.
CC   -!- SIMILARITY: Belongs to the Mo25 family.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AF151824; AAD34061.1; -.
DR   EMBL; AF113536; AAF14873.1; -.
DR   EMBL; BC020570; AAH20570.1; -.
DR   PDB; 1UPK; X-ray; A=1-341.
DR   IntAct; Q9Y376; -.
DR   Genew; HGNC:20292; CAB39.
DR   InterPro; IPR008938; ARM.
DR   InterPro; IPR004892; Mo25.
DR   Pfam; PF03204; Mo25; 1.
KW   3D-structure.
SQ   SEQUENCE   341 AA;  39869 MW;  EC710A528B6F9811 CRC64;

  Query Match             91.8%;  Score 1635.5;  DB 1;  Length 341;
  Best Local Similarity   94.0%;  Pred. No. 1e-96;
  Matches  329;  Conservative    1;  Mismatches   11;  Indels    9;  Gaps    2;

Qy          1 MPFPFGKSHKSPADIVKNLKESMAVLEKQDISDKKAEKATEEVSKNLVAMKEILYGTNEK 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MPFPFGKSHKSPADIVKNLKESMAVLEKQDISDKKAEKATEEVSKNLVAMKEILYGTNEK 60

Qy         61 EPQTEAVAQLAQELYNSGLLSTLVADLQLIDFEGKKDVAQIFNNILRRQIGTRTPTVEYI 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 EPQTEAVAQLAQELYNSGLLSTLVADLQLIDFEGKKDVAQIFNNILRRQIGTRTPTVEYI 120

Qy        121 CTQQNILFMLLKGYESPEIALNCGIMLRECIRHEPLAKIILWSEQFYDFFRYVEMSTFDI 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CTQQNILFMLLKGYESPEIALNCGIMLRECIRHEPLAKIILWSEQFYDFFRYVEMSTFDI 180

Qy        181 ASDAFATFKDLLTRHKLLSAEFLEQHYDRFFSEYEKLLHSENYVTKRQSLKLLGELLLDR 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 ASDAFATFKDLLTRHKLLSAEFLEQHYDRFFSEYEKLLHSENYVTKRQSLKLLGELLLDR 240

Qy        241 HNFTIMTKYISKPENLKLMMNLLRDKSRNIQFEAFHVFKVFVANPNKTQPILDILLKNQA 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 HNFTIMTKYISKPENLKLMMNLLRDKSRNIQFEAFHVFKVFVANPNKTQPILDILLKNQA 300

Qy        301 KLIEFLSKFQNDRTDCMSSSVPTTNSRVDLRVKPRTRGIRDLKRPAQQEA 350
              |||||||||||||:          |        ||      ||||||||||
Db        301 KLIEFLSKFQNDRTE-----DEQFNDEKTYLVKQ----IRDLKRPAQQEA 341


RESULT 2
Q8VDZ8
ID   Q8VDZ8      PRELIMINARY;      PRT;   341 AA.
AC   Q8VDZ8;
DT   01-MAR-2002 (TrEMBLrel. 20, Created)
DT   01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Cab39 protein.
GN   Name=Cab39;
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CZECH II;
RC   TISSUE=Mammary tumor metastatized to lung. Tumor arose spontaneously;
RX   MEDLINE=22388257; PubMed=12477932;
RA   Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA   Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA   Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA   Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA   Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA   Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA   Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA   Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
RA   Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA   Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA   Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA   Fahey J., Helton E., Ketteman M., Madan A., Rodrigues S., Sanchez A.,
RA   Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA   Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA   Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.,
RA   Krzywinski M.I., Skalska U., Smailus D.E., Schnerch A., Schein J.E.,
RA   Jones S.J., Marra M.A.;
RT   "Generation and initial analysis of more than 15,000 full-length human
RT   and mouse cDNA sequences.";
RL   Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CZECH II;
RC   TISSUE=Mammary tumor metastatized to lung. Tumor arose spontaneously;
RA   Strausberg R.;
RL   Submitted (DEC-2001) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; BC020041; AAH20041.1; -.
DR   MGD; MGI:107438; Cab39.
DR   InterPro; IPR008938; ARM.
DR   InterPro; IPR004892; Mo25.
DR   Pfam; PF03204; Mo25; 1.
SQ   SEQUENCE   341 AA;  39843 MW;  E7FECA529D6FE811 CRC64;

  Query Match             90.9%;  Score 1619.5;  DB 2;  Length 341;
  Best Local Similarity   93.1%;  Pred. No. 1.1e-95;
  Matches  326;  Conservative    1;  Mismatches   14;  Indels    9;  Gaps    2;

Qy          1 MPFPFGKSHKSPADIVKNLKESMAVLEKQDISDKKAEKATEEVSKNLVAMKEILYGTNEK 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MPFPFGKSHKSPADIVKNLKESMAVLEKQDISDKKAEKATEEVSKNLVAMKEILYGTNEK 60

Qy         61 EPQTEAVAQLAQELYNSGLLSTLVADLQLIDFEGKKDVAQIFNNILRRQIGTRTPTVEYI 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 EPQTEAVAQLAQELYNSGLLGTLVADLQLIDFEGKKDVAQIFNNILRRQIGTRTPTVEYI 120

Qy        121 CTQQNILFMLLKGYESPEIALNCGIMLRECIRHEPLAKIILWSEQFYDFFRYVEMSTFDI 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CTQQNILFMLLKGYESPEIALNCGIMLRECIRHEPLAKIILWSEQFYDFFRYVEMSTFDI 180

Qy        181 ASDAFATFKDLLTRHKLLSAEFLEQHYDRFFSEYEKLLHSENYVTKRQSLKLLGELLLDR 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 ASDAFATFKDLLTRHKLLSAEFLEQHYDRFFSEYEKLLHSENYVTKRQSLKLLGELLLDR 240

Qy        241 HNFTIMTKYISKPENLKLMMNLLRDKSRNIQFEAFHVFKVFVANPNKTQPILDILLKNQA 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 HNFTIMTKYISKPENLKLMMNLLRDKSRNIQFEAFHVFKVFVANPNKTQPILDILLKNQT 300

Qy        301 KLIEFLSKFQNDRTDCMSSSVPTTNSRVDLRVKPRTRGIRDLKRPAQQEA 350
              |||||||||||||:          |        ||      |||||| |||||
Db        301 KLIEFLSKFQNDRTE-----DEQFNDEKTYLVKQ----IRDLKRAAQQEA 341


RESULT 3
CB39_MOUSE
ID   CB39_MOUSE      STANDARD;      PRT;   341 AA.
AC   Q06138;
DT   01-FEB-1994 (Rel. 28, Created)
DT   01-FEB-1994 (Rel. 28, Last sequence update)
DT   01-OCT-2004 (Rel. 45, Last annotation update)
DE   Calcium binding protein 39 (Mo25 protein).
```

```
GN   Name=Cab39; Synonyms=Mo25;
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=93119656; PubMed=8418809;
RA   Miyamoto H., Matsushiro A., Nozaki M.;
RT   "Molecular cloning of a novel mRNA sequence expressed in cleavage
RT   stage mouse embryos.";
RL   Mol. Reprod. Dev. 34:1-7(1993).
CC   -!- FUNCTION: Together with the STE20-related adaptor-alpha (STRAD
CC       alpha) pseudo kinase, forms a regulatory complex capable of
CC       stimulating the activity of STK11.
CC   -!- SUBCELLULAR LOCATION: Cytoplasmic (Potential).
CC   -!- DEVELOPMENTAL STAGE: Transcribed during early mouse development.
CC       Detected at all developmental stages from the egg through the
CC       blastocyt, most abundant at the 2-cell stage.
CC   -!- SIMILARITY: Belongs to the Mo25 family.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; S51858; AAB24801.1; -.
DR   MGD; MGI:107438; Cab39.
DR   InterPro; IPR008938; ARM.
DR   InterPro; IPR004892; Mo25.
DR   Pfam; PF03204; Mo25; 1.
SQ   SEQUENCE   341 AA;  39842 MW;  E7F668529D6FE811 CRC64;

  Query Match             90.7%;  Score 1614.5;  DB 1;  Length 341;
  Best Local Similarity   92.9%;  Pred. No. 2.2e-95;
  Matches  325;  Conservative    2;  Mismatches   14;  Indels    9;  Gaps    2;

Qy          1 MPFPPFGKSHKSPADIVKNLKESMAVLEKQDISDKKAEKATEEVSKNLVAMKEILYGTNEK 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MPFPPFGKSHKSPADIVKNLKESMAVLEKQDISDKKAEKATEEVSKNLVAMKEILYGTNEK 60

Qy         61 EPQTEAVAQLAQELYNSGLLSTLVADLQLIDFEGKKDVAQIFNNILRRQIGTRTPTVEYI 120
              |||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||||
Db         61 EPQTEAVAQLAQELYNSGLLGTLVADLQLIDFEGKKDVAQIFNNILRRQIGTRTPTVEYI 120

Qy        121 CTQQNILFMLLKGYESPEIALNCGIMLRECIRHEPLAKIILWSEQFYDFFRYVEMSTFDI 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CTQQNILFMLLKGYESPEIALNCGIMLRECIRHEPLAKIILWSEQFYDFFRYVEMSTFDI 180

Qy        181 ASDAFATFKDLLTRHKLLSAEFLEQHYDRFFSEYEKLLHSENYVTKRQSLKLLGELLLDR 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 ASDAFATFKDLLTRHKLLSAEFLEQHYDRFFSEYEKLLHSENYVTKRQSLKLLGELLLDR 240

Qy        241 HNFTIMTKYISKPENLKLMMNLLRDKSRNIQFEAFHVFKVFVANPNKTQPILDILLKNQA 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 HNFTIMTKYISKPENLKLMMNLLRDKSRNIQFEAFHVFKVFVANPNKTQPILDILLKNQT 300

Qy        301 KLIEFLSKFQNDRTDCMSSSVPTTNSRVDLRVKPRTRGIRDLKRPAQQEA 350
              |||||||||||||:        |      ||     ||:||| |||||
Db        301 KLIEFLSKFQNDRTE-----DEQFNDEKTYLVKQ----IRNLKRAAQQEA 341


RESULT 4
Q6IP72
ID   Q6IP72      PRELIMINARY;      PRT;   341 AA.
AC   Q6IP72;
DT   05-JUL-2004 (TrEMBLrel. 27, Created)
DT   05-JUL-2004 (TrEMBLrel. 27, Last sequence update)
DT   05-JUL-2004 (TrEMBLrel. 27, Last annotation update)
DE   MGC78903 protein.
GN   Name=MGC78903;
OS   Xenopus laevis (African clawed frog).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidea; Pipidae;
OC   Xenopodinae; Xenopus.
OX   NCBI_TaxID=8355;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Spleen;
RX   MEDLINE=22388257; PubMed=12477932;
RA   Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA   Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA   Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA   Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA   Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA   Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA   Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA   Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
RA   Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA   Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA   Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA   Fahey J., Helton E., Ketteman M., Madan A., Rodrigues S., Sanchez A.,
RA   Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA   Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA   Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.,
RA   Krzywinski M.I., Skalska U., Smailus D.E., Schnerch A., Schein J.E.,
RA   Jones S.J., Marra M.A.;
RT   "Generation and initial analysis of more than 15,000 full-length human
RT   and mouse cDNA sequences.";
RL   Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Spleen;
RX   MEDLINE=22341132; PubMed=12454917;
RA   Klein S.L., Strausberg R.L., Wagner L., Pontius J., Clifton S.W.,
RA   Richardson P.;
RT   "Genetic and genomic tools for Xenopus research: The NIH Xenopus
RT   initiative.";
RL   Dev. Dyn. 225:384-391(2002).
RN   [3]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Spleen;
RA   Klein S., Strausberg R.;
RL   Submitted (JUN-2004) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; BC072045; AAH72045.1; -.
DR   InterPro; IPR008938; ARM.
DR   InterPro; IPR004892; Mo25.
DR   Pfam; PF03204; Mo25; 1.
SQ   SEQUENCE   341 AA;  39777 MW;  5AFEAF9325BB2F16 CRC64;

  Query Match             90.7%;  Score 1614.5;  DB 2;  Length 341;
  Best Local Similarity   92.3%;  Pred. No. 2.2e-95;
  Matches  323;  Conservative    5;  Mismatches   13;  Indels    9;  Gaps    2;

Qy          1 MPFPPFGKSHKSPADIVKNLKESMAVLEKQDISDKKAEKATEEVSKNLVAMKEILYGTNEK 60
              |||||||||||||||||||||:||||||||||||||||||||||||||||||||||||||
Db          1 MPFPPFGKSHKSPADIVKNLKESIAVLEKQDISDKKAEKATEEVSKNLVAMKEILYGTNEK 60

Qy         61 EPQTEAVAQLAQELYNSGLLSTLVADLQLIDFEGKKDVAQIFNNILRRQIGTRTPTVEYI 120
              |||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||||
Db         61 EPQTEAVAQLAQELYNSGLLGTLVADLQLIDFEGKKDVAQIFNNILRRQIGTRTPTVEYI 120

Qy        121 CTQQNILFMLLKGYESPEIALNCGIMLRECIRHEPLAKIILWSEQFYDFFRYVEMSTFDI 180
              ||||||||||||||||||||||||||||||||||||||||||| ||||||||||||||||
Db        121 CTQQNILFMLLKGYESPEIALNCGIMLRECIRHEPLAKIILWSEPFYDFFRYVEMSTFDI 180

Qy        181 ASDAFATFKDLLTRHKLLSAEFLEQHYDRFFSEYEKLLHSENYVTKRQSLKLLGELLLDR 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 ASDAFATFKDLLTRHKLLSAEFLEQHYDRFFSEYEKLLHSENYVTKRQSLKLLGELLLDR 240

Qy        241 HNFTIMTKYISKPENLKLMMNLLRDKSRNIQFEAFHVFKVFVANPNKTQPILDILLKNQA 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||:||||||:
Db        241 HNFTIMTKYISKPENLKLMMNLLRDKSRNIQFEAFHVFKVFVANPNKTQPVLDILLKNQS 300
```

```
Qy        301 KLIEFLSKFQNDRTDCMSSSVPTTNSRVDLRVKPRTRGIRDLKRPAQQEA 350
              ||||||||||||||:          |       ||      ||||||||||:|
Db        301 KLIEFLSKFQNDRTE-----DEQFNDEKTYLVKQ----IRDLKRPAQQDA 341
```

RESULT 5
AAH72045
ID AAH72045 PRELIMINARY; PRT; 341 AA.
AC AAH72045;
DT 01-JUN-2004 (TrEMBLrel. 27, Created)
DT 01-JUN-2004 (TrEMBLrel. 27, Last sequence update)
DT 01-JUN-2004 (TrEMBLrel. 27, Last annotation update)
DE Hypothetical protein.
OS Xenopus laevis (African clawed frog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidea; Pipidae;
OC Xenopodinae; Xenopus; Xenopus.
OX NCBI_TaxID=8355;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Spleen;
RX MEDLINE=22341132; PubMed=12454917;
RA Klein S.L., Strausberg R.L., Wagner L., Pontius J., Clifton S.W.,
RA Richardson P.;
RT "Genetic and genomic tools for Xenopus research: The NIH Xenopus
RT initiative.";
RL Dev. Dyn. 225:384-391(2002).
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=Spleen;
RX MEDLINE=22388257; PubMed=12477932;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahey J., Helton E., Ketteman M., Madan A., Rodrigues S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.,
RA Krzywinski M.I., Skalska U., Smailus D.E., Schnerch A., Schein J.E.,
RA Jones S.J., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length human
RT and mouse cDNA sequences.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN [3]
RP SEQUENCE FROM N.A.
RC TISSUE=Spleen;
RA Klein S., Strausberg R.;
RL Submitted (JUN-2004) to the EMBL/GenBank/DDBJ databases.
DR EMBL; BC072045; AAH72045.1; -.
KW Hypothetical protein.
SQ SEQUENCE 341 AA; 39777 MW; 5AFEAF9325BB2F16 CRC64;

Query Match 90.7%; Score 1614.5; DB 2; Length 341;
Best Local Similarity 92.3%; Pred. No. 2.2e-95;
Matches 323; Conservative 5; Mismatches 13; Indels 9; Gaps 2;

```
Qy          1 MPFPPFGKSHKSPADIVKNLKESMAVLEKQDISDKKAEKATEEVSKNLVAMKEILYGTNEK 60
              ||||||||||||||||||||:||||||||||||||||||||||||||||||||||||||||
Db          1 MPFPPFGKSHKSPADIVKNLKESIAVLEKQDISDKKAEKATEEVSKNLVAMKEILYGTNEK 60

Qy         61 EPQTEAVAQLAQELYNSGLLSTLVADLQLIDFEGKKDVAQIFNNILRRQIGTRTPTVEYI 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 EPQTEAVAQLAQELYNSGLLGTLVADLQLIDFEGKKDVAQIFNNILRRQIGTRTPTVEYI 120

Qy        121 CTQQNILFMLLKGYESPEIALNCGIMLRECIRHEPLAKIILWSEQFYDFFRYVEMSTFDI 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CTQQNILFMLLKGYESPEIALNCGIMLRECIRHEPLAKIILWSEPFYDFFRYVEMSTFDI 180

Qy        181 ASDAFATFKDLLTRHKLLSAEFLEQHYDRFFSEYEKLLHSENYVTKRQSLKLLGELLLDR 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 ASDAFATFKDLLTRHKLLSAEFLEQHYDRFFSEYEKLLHSENYVTKRQSLKLLGELLLDR 240

Qy        241 HNFTIMTKYISKPENLKLMMNLLRDKSRNIQFEAFHVFKVFVANPNKTQPILDILLKNQA 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||:||||||:
Db        241 HNFTIMTKYISKPENLKLMMNLLRDKSRNIQFEAFHVFKVFVANPNKTQPVLDILLKNQS 300

Qy        301 KLIEFLSKFQNDRTDCMSSSVPTTNSRVDLRVKPRTRGIRDLKRPAQQEA 350
              ||||||||||||||:          |       ||      ||||||||||:|
Db        301 KLIEFLSKFQNDRTE-----DEQFNDEKTYLVKQ----IRDLKRPAQQDA 341
```

RESULT 6
Q7Z4X0
ID Q7Z4X0 PRELIMINARY; PRT; 341 AA.
AC Q7Z4X0;
DT 01-OCT-2003 (TrEMBLrel. 25, Created)
DT 01-OCT-2003 (TrEMBLrel. 25, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE MO25-like protein.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Zhou Y., Yu L., Zhao S.Y.;
RL Submitted (MAR-1999) to the EMBL/GenBank/DDBJ databases.
DR EMBL; AF134480; AAP97257.1; -.
DR InterPro; IPR004892; Mo25.
DR Pfam; PF03204; Mo25; 1.
SQ SEQUENCE 341 AA; 39774 MW; D3A60160E78C7A6D CRC64;

Query Match 89.3%; Score 1590.5; DB 2; Length 341;
Best Local Similarity 92.3%; Pred. No. 7.6e-94;
Matches 323; Conservative 2; Mismatches 16; Indels 9; Gaps 2;

```
Qy          1 MPFPPFGKSHKSPADIVKNLKESMAVLEKQDISDKKAEKATEEVSKNLVAMKEILYGTNEK 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MPFPPFGKSHKSPADIVKNLKESMAVLEKQDISDKKAEKATEEVSKNLVAMKEILYGTNEK 60

Qy         61 EPQTEAVAQLAQELYNSGLLSTLVADLQLIDFEGKKDVAQIFNNILRRQIGTRTPTVEYI 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 EPQTEAVAQLAQELYNSGLLSTLVADLQLIDFEGKKDVAQIFNNILRRQIGTRTPTVEYI 120

Qy        121 CTQQNILFMLLKGYESPEIALNCGIMLRECIRHEPLAKIILWSEQFYDFFRYVEMSTFDI 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CTQQNILFMLLKGYESPEIALNCGIMLRECIRHEPLAKIILWSEQFYDFFRYVEMSTFDI 180

Qy        181 ASDAFATFKDLLTRHKLLSAEFLEQHYDRFFSEYEKLLHSENYVTKRQSLKLLGELLLDR 240
              ||||||||||||||||||||:     ||||||||||||||||||||||||||||||||||||
Db        181 ASDAFATFKDLLTRHKLLSAEFMVLLSFRFFSEYEKLLHSENYVTKRQSLKLLGELLLDR 240

Qy        241 HNFTIMTKYISKPENLKLMMNLLRDKSRNIQFEAFHVFKVFVANPNKTQPILDILLKNQA 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 HNFTIMTKYISKPENLKLMMNLLRDKSRNIQFEAFHVFKVFVANPNKTQPILDILLKNQA 300

Qy        301 KLIEFLSKFQNDRTDCMSSSVPTTNSRVDLRVKPRTRGIRDLKRPAQQEA 350
              ||||||||||||||:          |       ||      ||||||||||
Db        301 KLIEFLSKFQNDRTE-----DEQFNDEKTYLVKQ----IRDLKRPAQQEA 341
```

RESULT 7
Q6IQL2
ID Q6IQL2 PRELIMINARY; PRT; 341 AA.
AC Q6IQL2;
DT 05-JUL-2004 (TrEMBLrel. 27, Created)
DT 05-JUL-2004 (TrEMBLrel. 27, Last sequence update)

```
DT   05-JUL-2004 (TrEMBLrel. 27, Last annotation update)
DE   Zgc:86716.
GN   Name=zgc:86716;
OS   Brachydanio rerio (Zebrafish) (Danio rerio).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes;
OC   Cyprinidae; Danio.
OX   NCBI_TaxID=7955;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Embryo;
RX   MEDLINE=22388257; PubMed=12477932;
RA   Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA   Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA   Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA   Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA   Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA   Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA   Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA   Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
RA   Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA   Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA   Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA   Fahey J., Helton E., Ketteman M., Madan A., Rodrigues S., Sanchez A.,
RA   Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA   Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA   Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.,
RA   Krzywinski M.I., Skalska U., Smailus D.E., Schnerch A., Schein J.E.,
RA   Jones S.J., Marra M.A.;
RT   "Generation and initial analysis of more than 15,000 full-length human
RT   and mouse cDNA sequences.";
RL   Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Embryo;
RA   Strausberg R.;
RL   Submitted (JUN-2004) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; BC071393; AAH71393.1; -.
DR   InterPro; IPR008938; ARM.
DR   InterPro; IPR004892; Mo25.
DR   Pfam; PF03204; Mo25; 1.
SQ   SEQUENCE   341 AA;  39792 MW;  9CEE4A665F3C6614 CRC64;

  Query Match             86.3%;  Score 1537.5;  DB 2;  Length 341;
  Best Local Similarity   87.4%;  Pred. No. 1.9e-90;
  Matches  306;  Conservative   14;  Mismatches   21;  Indels    9;  Gaps    1;

Qy          1 MPFPFGKSHKSPADIVKNLKESMAVLEKQDISDKKAEKATEEVSKNLVAMKEILYGTNEK 60
              ||||| |||| |||||||||::| :|||||||||||||:|||||:|::||||||||||
Db          1 MPFPFVKSHKCPADIVKNLKDNMTILEKQDISDKKAEKASEEVSKSLLSMKEILYGTNEK 60

Qy         61 EPQTEAVAQLAQELYNSGLLSTLVADLQLIDFEGKKDVAQIFNNILRRQIGTRTPTVEYI 120
              ||||||||||||||||||||||||||||||||||||||||:|||||||||||||||:
Db         61 EPQTEAVAQLAQELYNSGLLSTLVADLQLIDFEGKKDVAQIFNDILRRQIGTRTPTVEYL 120

Qy        121 CTQQNILFMLLKGYESPEIALNCGIMLRECIRHEPLAKIILWSEQFYDFFRYVEMSTFDI 180
              ||||||||||||||||:|||||||||||||||||||||| | ||||||||||||||||||
Db        121 CTQQNILFMLLKGYESPDIALNCGIMLRECIRHEPLAKITLCSEQFYDFFRYVEMSTFDI 180

Qy        181 ASDAFATFKDLLTRHKLLSAEFLEQHYDRFFSEYEKLLHSENYVTKRQSLKLLGELLLDR 240
              |||||||||||||||:||||||||||||||||||||||||||||||||||||||||||||
Db        181 ASDAFATFKDLLTRHKVLSAEFLEQHYDRFFSEYEKLLHSENYVTKRQSLKLLGELLLDR 240

Qy        241 HNFTIMTKYISKPENLKLMMNLLRDKSRNIQFEAFHVFKVFVANPNKTQPILDILLKNQA 300
              |||||||||||||||||||||||||||||||||||||||||||||||| ||||||||||
Db        241 HNFTIMTKYISKPENLKLMMNLLRDKSRNIQFEAFHVFKVFVANPNKTPPILDILLKNQT 300

Qy        301 KLIEFLSKFQNDRTDCMSSSVPTTNSRVDLRVKPRTRGIRDLKRPAQQEA 350
              |||||||||||| :    |   |           : |||||||  |:|
Db        301 KLIEFLSKFQNDRAEDEQFSDEKTY---------LIKQIRDLKRPTPQDA 341


RESULT 8
AAH71393
ID   AAH71393    PRELIMINARY;      PRT;   341 AA.
AC   AAH71393;
DT   01-JUN-2004 (TrEMBLrel. 27, Created)
DT   01-JUN-2004 (TrEMBLrel. 27, Last sequence update)
DT   01-JUN-2004 (TrEMBLrel. 27, Last annotation update)
DE   Hypothetical protein.
OS   Brachydanio rerio (Zebrafish) (Danio rerio).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes;
OC   Cyprinidae; Danio.
OX   NCBI_TaxID=7955;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Embryo;
RX   MEDLINE=22388257; PubMed=12477932;
RA   Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA   Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA   Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA   Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA   Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA   Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA   Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA   Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
RA   Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA   Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA   Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA   Fahey J., Helton E., Ketteman M., Madan A., Rodrigues S., Sanchez A.,
RA   Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA   Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA   Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.,
RA   Krzywinski M.I., Skalska U., Smailus D.E., Schnerch A., Schein J.E.,
RA   Jones S.J., Marra M.A.;
RT   "Generation and initial analysis of more than 15,000 full-length human
RT   and mouse cDNA sequences.";
RL   Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Embryo;
RA   Strausberg R.;
RL   Submitted (JUN-2004) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; BC071393; AAH71393.1; -.
KW   Hypothetical protein.
SQ   SEQUENCE   341 AA;  39792 MW;  9CEE4A665F3C6614 CRC64;

  Query Match             86.3%;  Score 1537.5;  DB 2;  Length 341;
  Best Local Similarity   87.4%;  Pred. No. 1.9e-90;
  Matches  306;  Conservative   14;  Mismatches   21;  Indels    9;  Gaps    1;

Qy          1 MPFPFGKSHKSPADIVKNLKESMAVLEKQDISDKKAEKATEEVSKNLVAMKEILYGTNEK 60
              ||||| |||| |||||||||::| :|||||||||||||:|||||:|::||||||||||
Db          1 MPFPFVKSHKCPADIVKNLKDNMTILEKQDISDKKAEKASEEVSKSLLSMKEILYGTNEK 60

Qy         61 EPQTEAVAQLAQELYNSGLLSTLVADLQLIDFEGKKDVAQIFNNILRRQIGTRTPTVEYI 120
              ||||||||||||||||||||||||||||||||||||||||:|||||||||||||||:
Db         61 EPQTEAVAQLAQELYNSGLLSTLVADLQLIDFEGKKDVAQIFNDILRRQIGTRTPTVEYL 120

Qy        121 CTQQNILFMLLKGYESPEIALNCGIMLRECIRHEPLAKIILWSEQFYDFFRYVEMSTFDI 180
              ||||||||||||||||:|||||||||||||||||||||| | ||||||||||||||||||
Db        121 CTQQNILFMLLKGYESPDIALNCGIMLRECIRHEPLAKITLCSEQFYDFFRYVEMSTFDI 180

Qy        181 ASDAFATFKDLLTRHKLLSAEFLEQHYDRFFSEYEKLLHSENYVTKRQSLKLLGELLLDR 240
              |||||||||||||||:||||||||||||||||||||||||||||||||||||||||||||
Db        181 ASDAFATFKDLLTRHKVLSAEFLEQHYDRFFSEYEKLLHSENYVTKRQSLKLLGELLLDR 240

Qy        241 HNFTIMTKYISKPENLKLMMNLLRDKSRNIQFEAFHVFKVFVANPNKTQPILDILLKNQA 300
              |||||||||||||||||||||||||||||||||||||||||||||||| ||||||||||
Db        241 HNFTIMTKYISKPENLKLMMNLLRDKSRNIQFEAFHVFKVFVANPNKTPPILDILLKNQT 300

Qy        301 KLIEFLSKFQNDRTDCMSSSVPTTNSRVDLRVKPRTRGIRDLKRPAQQEA 350
              |||||||||||| :    |   |           : |||||||  |:|
Db        301 KLIEFLSKFQNDRAEDEQFSDEKTY---------LIKQIRDLKRPTPQDA 341
```

```
RESULT 9
Q6PAB4
ID   Q6PAB4      PRELIMINARY;      PRT;   337 AA.
AC   Q6PAB4;
DT   05-JUL-2004 (TrEMBLrel. 27, Created)
DT   05-JUL-2004 (TrEMBLrel. 27, Last sequence update)
DT   05-JUL-2004 (TrEMBLrel. 27, Last annotation update)
DE   MGC68674 protein.
GN   Name=MGC68674;
OS   Xenopus laevis (African clawed frog).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidea; Pipidae;
OC   Xenopodinae; Xenopus.
OX   NCBI_TaxID=8355;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Kidney;
RX   MEDLINE=22388257; PubMed=12477932;
RA   Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA   Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA   Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA   Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA   Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA   Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA   Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA   Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
RA   Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA   Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA   Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA   Fahey J., Helton E., Ketteman M., Madan A., Rodrigues S., Sanchez A.,
RA   Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA   Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA   Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.,
RA   Krzywinski M.I., Skalska U., Smailus D.E., Schnerch A., Schein J.E.,
RA   Jones S.J., Marra M.A.;
RT   "Generation and initial analysis of more than 15,000 full-length human
RT   and mouse cDNA sequences.";
RL   Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Kidney;
RX   MEDLINE=22341132; PubMed=12454917;
RA   Klein S.L., Strausberg R.L., Wagner L., Pontius J., Clifton S.W.,
RA   Richardson P.;
RT   "Genetic and genomic tools for Xenopus research: The NIH Xenopus
RT   initiative.";
RL   Dev. Dyn. 225:384-391(2002).
RN   [3]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Kidney;
RA   Klein S., Strausberg R.;
RL   Submitted (OCT-2003) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; BC060384; AAH60384.1; -.
DR   InterPro; IPR008938; ARM.
DR   InterPro; IPR004892; Mo25.
DR   Pfam; PF03204; Mo25; 1.
SQ   SEQUENCE   337 AA;  39183 MW;  CCAE68F5D8569B57 CRC64;

  Query Match             75.2%;  Score 1339.5;  DB 2;  Length 337;
  Best Local Similarity   77.7%;  Pred. No. 8.5e-78;
  Matches  268;  Conservative   35;  Mismatches   29;  Indels   13;  Gaps    4;

Qy          1 MPFPFGKSHKSPADIVKNLKESMAVLEKQDISDKKAEKATEEVSKNLVAMKEILYGTNEK 60
              ||  | ||||:||:||| ||::||:||:|   ||| |||:|||||:| | |||| || :|
Db          4 MPL-FSKSHKNPAEIVKTLKDNMALLERQ---DKKTEKASEEVSKSLQATKEILCGTGDK 59

Qy         61 EPQTEAVAQLAQELYNSGLLSTLVADLQLIDFEGKKDVAQIFNNILRRQIGTRTPTVEYI 120
              ||||| |||||||||||||| ||:|:| |||||||||||:|||||||||||||:||||||
Db         60 EPQTETVAQLAQELYNSGLLVTLIANLHLIDFEGKKDVSQIFNNILRRQIGTRSPTVEYI 119

Qy        121 CTQQNILFMLLKGYESPEIALNCGIMLRECIRHEPLAKIILWSEQFYDFFRYVEMSTFDI 180
                : |:|||:|||||||||::||:||||||||:|||||||:||:|||| ||||:|||||||||

Db        120 SSHQHILFILLKGYESPQVALHCGIMLRECVRHEPLAKVILYSEQFGDFFKYVEMSTFDI 179

Qy        181 ASDAFATFKDLLTRHKLLSAEFLEQHYDRFFSEYEKLLHSENYVTKRQSLKLLGELLLDR 240
              |||||||||||||||||: |||||:||| |::|||||||||||||||||||||:|||
Db        180 ASDAFATFKDLLTRHKLMVAEFLEQNYDRIFNDYEKLLHSENYVTKRQSLKLLGELILDR 239

Qy        241 HNFTIMTKYISKPENLKLMMNLLRDKSRNIQFEAFHVFKVFVANPNKTQPILDILLKNQA 300
              |||:|||||||||||||||||||| |||||||||||||||||||:|||||||
Db        240 HNFSIMTKYISKPENLKLMMNLLRDKSPNIQFEAFHVFKVFVANPNKTQPIVDILLKNQT 299

Qy        301 KLIEFLSKFQNDRTDCMSSSVPTTNSRVDLRVKPRTRGIRDLKRP 345
              |||:||| || ||||       |: :  |      : |||||:|
Db        300 KLIDFLSSFQKDRTD----DEQFTDEKNYL-----IKQIRDLKKP 335


RESULT 10
AAH60384
ID   AAH60384    PRELIMINARY;      PRT;   337 AA.
AC   AAH60384;
DT   02-MAR-2004 (TrEMBLrel. 27, Created)
DT   02-MAR-2004 (TrEMBLrel. 27, Last sequence update)
DT   02-MAR-2004 (TrEMBLrel. 27, Last annotation update)
DE   MGC68674 protein.
GN   MGC68674.
OS   Xenopus laevis (African clawed frog).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidea; Pipidae;
OC   Xenopodinae; Xenopus; Xenopus.
OX   NCBI_TaxID=8355;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Kidney;
RX   MEDLINE=22341132; PubMed=12454917;
RA   Klein S.L., Strausberg R.L., Wagner L., Pontius J., Clifton S.W.,
RA   Richardson P.;
RT   "Genetic and genomic tools for Xenopus research: The NIH Xenopus
RT   initiative.";
RL   Dev. Dyn. 225:384-391(2002).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Kidney;
RX   MEDLINE=22388257; PubMed=12477932;
RA   Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA   Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA   Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA   Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA   Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA   Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA   Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA   Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
RA   Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA   Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA   Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA   Fahey J., Helton E., Ketteman M., Madan A., Rodrigues S., Sanchez A.,
RA   Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA   Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA   Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.,
RA   Krzywinski M.I., Skalska U., Smailus D.E., Schnerch A., Schein J.E.,
RA   Jones S.J., Marra M.A.;
RT   "Generation and initial analysis of more than 15,000 full-length human
RT   and mouse cDNA sequences.";
RL   Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN   [3]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Kidney;
RA   Klein S., Strausberg R.;
RL   Submitted (OCT-2003) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; BC060384; AAH60384.1; -.
SQ   SEQUENCE   337 AA;  39183 MW;  CCAE68F5D8569B57 CRC64;

  Query Match             75.2%;  Score 1339.5;  DB 2;  Length 337;
  Best Local Similarity   77.7%;  Pred. No. 8.5e-78;
  Matches  268;  Conservative   35;  Mismatches   29;  Indels   13;  Gaps    4;
```

```
Qy          1 MPFPFGKSHKSPADIVKNLKESMAVLEKQDISDKKAEKATEEVSKNLVAMKEILYGTNEK 60
              ||  | ||||:||:||| ||::||:||:|   ||| |||:|||||:| | |||| || :|
Db          4 MPL-FSKSHKNPAEIVKTLKDNMALLERQ---DKKTEKASEEVSKSLQATKEILCGTGDK 59

Qy         61 EPQTEAVAQLAQELYNSGLLSTLVADLQLIDFEGKKDVAQIFNNILRRQIGTRTPTVEYI 120
              ||||| |||||||||||||| ||:|:| ||||||||||:|||||||||||||:|||||
Db         60 EPQTETVAQLAQELYNSGLLVTLIANLHLIDFEGKKDVSQIFNNILRRQIGTRSPTVEYI 119

Qy        121 CTQQNILFMLLKGYESPEIALNCGIMLRECIRHEPLAKIILWSEQFYDFFRYVEMSTFDI 180
              : |:|||:||||||||::||:||||||||:|||||||:||:|||| |||:|||||||||
Db        120 SSHQHILFILLKGYESPQVALHCGIMLRECVRHEPLAKVILYSEQFGDFFKYVEMSTFDI 179

Qy        181 ASDAFATFKDLLTRHKLLSAEFLEQHYDRFFSEYEKLLHSENYVTKRQSLKLLGELLLDR 240
              ||||||||||||||||: ||||||:||| |::|||||||||||||||||||||:|||
Db        180 ASDAFATFKDLLTRHKLMVAEFLEQNYDRIFNDYEKLLHSENYVTKRQSLKLLGELILDR 239

Qy        241 HNFTIMTKYISKPENLKLMMNLLRDKSRNIQFEAFHVFKVFVANPNKTQPILDILLKNQA 300
              |||:|||||||||||||||||||||| ||||||||||||||||||:|||||||
Db        240 HNFSIMTKYISKPENLKLMMNLLRDKSPNIQFEAFHVFKVFVANPNKTQPIVDILLKNQT 299

Qy        301 KLIEFLSKFQNDRTDCMSSSVPTTNSRVDLRVKPRTRGIRDLKRP 345
              |||:||| || |||| |: :  |  : |||||:|
Db        300 KLIDFLSSFQKDRTD----DEQFTDEKNYL-----IKQIRDLKKP 335
```

RESULT 11
C39L_MOUSE
ID C39L_MOUSE STANDARD; PRT; 337 AA.
AC Q9DB16; Q8BG52; Q91WB8; Q91YL0;
DT 16-OCT-2001 (Rel. 40, Created)
DT 01-OCT-2004 (Rel. 45, Last sequence update)
DT 01-OCT-2004 (Rel. 45, Last annotation update)
DE Calcium binding protein 39-like (Mo25-like protein).
GN Name=Cab391;
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A. (ISOFORMS 1 AND 2).
RC STRAIN=C57BL/6J; TISSUE=Cerebellum, Eye, Pituitary, and Testis;
RX MEDLINE=22354683; PubMed=12466851; DOI=10.1038/nature01266;
RA Okazaki Y., Furuno M., Kasukawa T., Adachi J., Bono H., Kondo S.,
RA Nikaido I., Osato N., Saito R., Suzuki H., Yamanaka I., Kiyosawa H.,
RA Yagi K., Tomaru Y., Hasegawa Y., Nogami A., Schonbach C., Gojobori T.,
RA Baldarelli R., Hill D.P., Bult C., Hume D.A., Quackenbush J.,
RA Schriml L.M., Kanapin A., Matsuda H., Batalov S., Beisel K.W.,
RA Blake J.A., Bradt D., Brusic V., Chothia C., Corbani L.E., Cousins S.,
RA Dalla E., Dragani T.A., Fletcher C.F., Forrest A., Frazer K.S.,
RA Gaasterland T., Gariboldi M., Gissi C., Godzik A., Gough J.,
RA Grimmond S., Gustincich S., Hirokawa N., Jackson I.J., Jarvis E.D.,
RA Kanai A., Kawaji H., Kawasawa Y., Kedzierski R.M., King B.L.,
RA Konagaya A., Kurochkin I.V., Lee Y., Lenhard B., Lyons P.A.,
RA Maglott D.R., Maltais L., Marchionni L., McKenzie L., Miki H.,
RA Nagashima T., Numata K., Okido T., Pavan W.J., Pertea G., Pesole G.,
RA Petrovsky N., Pillai R., Pontius J.U., Qi D., Ramachandran S.,
RA Ravasi T., Reed J.C., Reed D.J., Reid J., Ring B.Z., Ringwald M.,
RA Sandelin A., Schneider C., Semple C.A., Setou M., Shimada K.,
RA Sultana R., Takenaka Y., Taylor M.S., Teasdale R.D., Tomita M.,
RA Verardo R., Wagner L., Wahlestedt C., Wang Y., Watanabe Y., Wells C.,
RA Wilming L.G., Wynshaw-Boris A., Yanagisawa M., Yang I., Yang L.,
RA Yuan Z., Zavolan M., Zhu Y., Zimmer A., Carninci P., Hayatsu N.,
RA Hirozane-Kishikawa T., Konno H., Nakamura M., Sakazume N., Sato K.,
RA Shiraki T., Waki K., Kawai J., Aizawa K., Arakawa T., Fukuda S.,
RA Hara A., Hashizume W., Imotani K., Ishii Y., Itoh M., Kagawa I.,
RA Miyazaki A., Sakai K., Sasaki D., Shibata K., Shinagawa A.,
RA Yasunishi A., Yoshino M., Waterston R., Lander E.S., Rogers J.,
RA Birney E., Hayashizaki Y.;
RT "Analysis of the mouse transcriptome based on functional annotation of
RT 60,770 full-length cDNAs.";
RL Nature 420:563-573(2002).
RN [2]
RP SEQUENCE FROM N.A. (ISOFORM 1).
RC STRAIN=FVB/N; TISSUE=Mammary gland, and Salivary gland;
RX MEDLINE=22388257; PubMed=12477932; DOI=10.1073/pnas.242603899;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahey J., Helton E., Ketteman M., Madan A., Rodrigues S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smailus D.E.,
RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length human
RT and mouse cDNA sequences.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
CC -!- ALTERNATIVE PRODUCTS:
CC   Event=Alternative splicing; Named isoforms=2;
CC   Name=1;
CC   IsoId=Q9DB16-1; Sequence=Displayed;
CC   Name=2;
CC   IsoId=Q9DB16-2; Sequence=VSP_007417, VSP_007418;
CC   Note=No experimental confirmation available;
CC -!- SIMILARITY: Belongs to the Mo25 family.
CC ---
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC ---
DR EMBL; AK005323; BAB23953.2; -.
DR EMBL; AK030474; BAC26978.1; -.
DR EMBL; AK053642; BAC35457.1; -.
DR EMBL; AK076758; BAC36470.1; -.
DR EMBL; AK076867; BAC36513.1; -.
DR EMBL; BC016128; AAH16128.1; ALT_INIT.
DR EMBL; BC016546; AAH16546.2; -.
DR MGD; MGI:1914081; 4930520C08Rik.
DR InterPro; IPR008938; ARM.
DR InterPro; IPR004892; Mo25.
DR Pfam; PF03204; Mo25; 1.
KW Alternative splicing.
FT VARSPLIC 279 296 VFVASPHKTQPIVEILLK -> NSVFITNRIHGLKRWLSS
FT   (in isoform 2).
FT   /FTId=VSP_007417.
FT VARSPLIC 297 337 Missing (in isoform 2).
FT   /FTId=VSP_007418.
FT CONFLICT   45   45 S -> P (in Ref. 1; BAB23953).
FT CONFLICT 232 232 L -> R (in Ref. 2; AAH16546).
SQ SEQUENCE 337 AA; 39105 MW; C62B5B58095A98C8 CRC64;

```
  Query Match             72.9%;  Score 1298.5;  DB 1;  Length 337;
  Best Local Similarity   76.0%;  Pred. No. 3.6e-75;
  Matches  263;  Conservative   33;  Mismatches   37;  Indels   13;  Gaps    3;

Qy          1 MPFPFGKSHKSPADIVKNLKESMAVLEKQDISDKKAEKATEEVSKNLVAMKEILYGTNEK 60
              ||  | ||||:||:||| ||:::|:||||   ||| :||:|||||:| |||||| |||:|
Db          4 MPL-FSKSHKNPAEIVKILKDNLAILEKQ---DKKTDKASEEVSKSLQAMKEILCGTNDK 59

Qy         61 EPQTEAVAQLAQELYNSGLLSTLVADLQLIDFEGKKDVAQIFNNILRRQIGTRTPTVEYI 120
              || |||||||||||:|||| ||:|||||||||||||||| ||||||||||||| ||||||
Db         60 EPPTEAVAQLAQELYSSGLLVTLIADLQLIDFEGKKDVTQIFNNILRRQIGTRCPTVEYI 119
```

```
Qy        121 CTQQNILFMLLKGYESPEIALNCGIMLRECIRHEPLAKIILWSEQFYDFFRYVEMSTFDI 180
              :  :|||||||||||:|:||| ||||||||||||||||||:| || |||:|||:|||||
Db        120 SSHPHILFMLLKGYEAPQIALRCGIMLRECIRHEPLAKIILFSNQFRDFFKYVELSTFDI 179

Qy        181 ASDAFATFKDLLTRHKLLSAEFLEQHYDRFFSEYEKLLHSENYVTKRQSLKLLGELLLDR 240
              |||||||||||||||:| |:||||:||  | :||||| ||||||||||||||||:|||
Db        180 ASDAFATFKDLLTRHKVLVADFLEQNYDTIFEDYEKLLQSENYVTKRQSLKLLGELILDR 239

Qy        241 HNFTIMTKYISKPENLKLMMNLLRDKSRNIQFEAFHVFKVFVANPNKTQPILDILLKNQA 300
              |||||||||||||||||||||||| |||||||||||||:|:|||||::||||||
Db        240 HNFTIMTKYISKPENLKLMMNLLRDKSPNIQFEAFHVFKVFVASPHKTQPIVEILLKNQP 299

Qy        301 KLIEFLSKFQNDRTDCMSSSVPTTNSRVDLRVKPRTRGIRDLKRPA 346
              ||||||| || :|||         : :         : |||||: |
Db        300 KLIEFLSSFQKERTD---------DEQFADEKNYLIKQIRDLKKAA 336
```

RESULT 12
C39L_HUMAN
ID C39L_HUMAN STANDARD; PRT; 337 AA.
AC Q9H9S4; Q6WG71; Q96FG1; Q9BZ33;
DT 16-OCT-2001 (Rel. 40, Created)
DT 01-OCT-2004 (Rel. 45, Last sequence update)
DT 01-OCT-2004 (Rel. 45, Last annotation update)
DE Calcium binding protein 39-like (Mo25-like protein) (Antigen MLAA-34).
GN Name=CAB39L;
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Chen G., Zhang W.G., Fu J., Cao X.M., Chen Y.X., Zhao W.H., Zhao A.Z.,
RA Han Y.H., Li F.Y., Liu X.P., Yao L.B.;
RT "Serological identification of immunogenic antigens in acute monocytic
RT leukemia.";
RL Submitted (MAY-2003) to the EMBL/GenBank/DDBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=Placenta;
RX MEDLINE=22388257; PubMed=12477932; DOI=10.1073/pnas.242603899;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahey J., Helton E., Ketteman M., Madan A., Rodrigues S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smailus D.E.,
RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length human
RT and mouse cDNA sequences.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN [3]
RP SEQUENCE FROM N.A.
RX PubMed=15057823; DOI=10.1038/nature02379;
RA Dunham A., Matthews L.H., Burton J., Ashurst J.L., Howe K.L.,
RA Ashcroft K.J., Beare D.M., Burford D.C., Hunt S.E.,
RA Griffiths-Jones S., Jones M.C., Keenan S.J., Oliver K., Scott C.E.,
RA Ainscough R., Almeida J.P., Ambrose K.D., Andrews D.T.,
RA Ashwell R.I.S., Babbage A.K., Bagguley C.L., Bailey J., Bannerjee R.,
RA Barlow K.F., Bates K., Beasley H., Bird C.P., Bray-Allen S.,
RA Brown A.J., Brown J.Y., Burrill W., Carder C., Carter N.P.,
RA Chapman J.C., Clamp M.E., Clark S.Y., Clarke G., Clee C.M.,
RA Clegg S.C., Cobley V., Collins J.E., Corby N., Coville G.J.,
RA Deloukas P., Dhami P., Dunham I., Dunn M., Earthrowl M.E.,
RA Ellington A.G., Faulkner L., Frankish A.G., Frankland J., French L.,
RA Garner P., Garnett J., Gilbert J.G.R., Gilson C.J., Ghori J.,
RA Grafham D.V., Gribble S.M., Griffiths C., Hall R.E., Hammond S.,
RA Harley J.L., Hart E.A., Heath P.D., Howden P.J., Huckle E.J.,
RA Hunt P.J., Hunt A.R., Johnson C., Johnson D., Kay M., Kimberley A.M.,
RA King A., Laird G.K., Langford C.J., Lawlor S., Leongamornlert D.A.,
RA Lloyd D.M., Lloyd C., Loveland J.E., Lovell J., Martin S.,
RA Mashreghi-Mohammadi M., McLaren S.J., McMurray A., Milne S.,
RA Moore M.J.F., Nickerson T., Palmer S.A., Pearce A.V., Peck A.I.,
RA Pelan S., Phillimore B., Porter K.M., Rice C.M., Searle S.,
RA Sehra H.K., Shownkeen R., Skuce C.D., Smith M., Steward C.A.,
RA Sycamore N., Tester J., Thomas D.W., Tracey A., Tromans A., Tubby B.,
RA Wall M., Wallis J.M., West A.P., Whitehead S.L., Willey D.L.,
RA Wilming L., Wray P.W., Wright M.W., Young L., Coulson A., Durbin R.,
RA Hubbard T., Sulston J.E., Beck S., Bentley D.R., Rogers J., Ross M.T.;
RT "The DNA sequence and analysis of human chromosome 13.";
RL Nature 428:522-528(2004).
RN [4]
RP SEQUENCE OF 7-337 FROM N.A.
RX PubMed=14702039; DOI=10.1038/ng1285;
RA Ota T., Suzuki Y., Nishikawa T., Otsuki T., Sugiyama T., Irie R.,
RA Wakamatsu A., Hayashi K., Sato H., Nagai K., Kimura K., Makita H.,
RA Sekine M., Obayashi M., Nishi T., Shibahara T., Tanaka T., Ishii S.,
RA Yamamoto J.-I., Saito K., Kawai Y., Isono Y., Nakamura Y.,
RA Nagahari K., Murakami K., Yasuda T., Iwayanagi T., Wagatsuma M.,
RA Shiratori A., Sudo H., Hosoiri T., Kaku Y., Kodaira H., Kondo H.,
RA Sugawara M., Takahashi M., Kanda K., Yokoi T., Furuya T., Kikkawa E.,
RA Omura Y., Abe K., Kamihara K., Katsuta N., Sato K., Tanikawa M.,
RA Yamazaki M., Ninomiya K., Ishibashi T., Yamashita H., Murakawa K.,
RA Fujimori K., Tanai H., Kimata M., Watanabe M., Hiraoka S., Chiba Y.,
RA Ishida S., Ono Y., Takiguchi S., Watanabe S., Yosida M., Hotuta T.,
RA Kusano J., Kanehori K., Takahashi-Fujii A., Hara H., Tanase T.-O.,
RA Nomura Y., Togiya S., Komai F., Hara R., Takeuchi K., Arita M.,
RA Imose N., Musashino K., Yuuki H., Oshima A., Sasaki N., Aotsuka S.,
RA Yoshikawa Y., Matsunawa H., Ichihara T., Shiohata N., Sano S.,
RA Moriya S., Momiyama H., Satoh N., Takami S., Terashima Y., Suzuki O.,
RA Nakagawa S., Senoh A., Mizoguchi H., Goto Y., Shimizu F., Wakebe H.,
RA Hishigaki H., Watanabe T., Sugiyama A., Takemoto M., Kawakami B.,
RA Yamazaki M., Watanabe K., Kumagai A., Itakura S., Fukuzumi Y.,
RA Fujimori Y., Komiyama M., Tashiro H., Tanigami A., Fujiwara T.,
RA Ono T., Yamada K., Fujii Y., Ozaki K., Hirao M., Ohmori Y.,
RA Kawabata A., Hikiji T., Kobatake N., Inagaki H., Ikema Y., Okamoto S.,
RA Okitani R., Kawakami T., Noguchi S., Itoh T., Shigeta K., Senba T.,
RA Matsumura K., Nakajima Y., Mizuno T., Morinaga M., Sasaki M.,
RA Togashi T., Oyama M., Hata H., Watanabe M., Komatsu T.,
RA Mizushima-Sugano J., Satoh T., Shirai Y., Takahashi Y., Nakagawa K.,
RA Okumura K., Nagase T., Nomura N., Kikuchi H., Masuho Y., Yamashita R.,
RA Nakai K., Yada T., Nakamura Y., Ohara O., Isogai T., Sugano S.;
RT "Complete sequencing and characterization of 21,243 full-length human
RT cDNAs.";
RL Nat. Genet. 36:40-45(2004).
CC -!- SIMILARITY: Belongs to the Mo25 family.
CC ---
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC ---
DR EMBL; AY288977; AAQ93064.1; -.
DR EMBL; BC010993; AAH10993.2; -.
DR EMBL; AL138875; CAC28084.1; -.
DR EMBL; AK022639; BAB14147.1; ALT_INIT.
DR Genew; HGNC:20290; CAB39L.
DR InterPro; IPR008938; ARM.
DR InterPro; IPR004892; Mo25.
DR Pfam; PF03204; Mo25; 1.
FT CONFLICT 58 58 E -> D (in Ref. 1).
FT CONFLICT 93 93 G -> E (in Ref. 2).
FT CONFLICT 113 113 S -> C (in Ref. 1).

```
FT   CONFLICT    116    116       V -> A (in Ref. 1).
FT   CONFLICT    121    121       A -> S (in Ref. 1).
FT   CONFLICT    127    127       F -> V (in Ref. 1).
FT   CONFLICT    156    156       A -> V (in Ref. 4).
FT   CONFLICT    243    243       A -> T (in Ref. 1).
FT   CONFLICT    335    335       T -> A (in Ref. 1).
SQ   SEQUENCE   337 AA;  39087 MW;  5C6C7A83B9B4CA2D CRC64;

  Query Match             72.9%;  Score 1297.5;  DB 1;  Length 337;
  Best Local Similarity   76.0%;  Pred. No. 4.1e-75;
  Matches  263;  Conservative   32;  Mismatches   38;  Indels   13;  Gaps    3;

Qy          1 MPFPPFGKSHKSPADIVKNLKESMAVLEKQDISDKKAEKATEEVSKNLVAMKEILYGTNEK 60
              ||  | ||||:||:||| ||:::|:||||   ||| :||:|||||:| |||||| |||||
Db          4 MPL-FSKSHKNPAEIVKILKDNLAILEKQ---DKKTDKASEEVSKSLQAMKEILCGTNEK 59

Qy         61 EPQTEAVAQLAQELYNSGLLSTLVADLQLIDFEGKKDVAQIFNNILRRQIGTRTPTVEYI 120
              || ||||||||||||:|||| ||:|||||||||||||| |||||||||||||:||||||
Db         60 EPPTEAVAQLAQELYSSGLLVTLIADLQLIDFEGKKDVTQIFNNILRRQIGTRSPTVEYI 119

Qy        121 CTQQNILFMLLKGYESPEIALNCGIMLRECIRHEPLAKIILWSEQFYDFFRYVEMSTFDI 180
                :|||||||||:|:||| |||||||||||||||||:| || |||:|||:|||||
Db        120 SAHPHILFMLLKGYEAPQIALRCGIMLRECIRHEPLAKIILFSNQFRDFFKYVELSTFDI 179

Qy        181 ASDAFATFKDLLTRHKLLSAEFLEQHYDRFFSEYEKLLHSENYVTKRQSLKLLGELLLDR 240
              |||||||||||||||:| |:||||:||  | :||||| |||||||||||||||:|||
Db        180 ASDAFATFKDLLTRHKVLVADFLEQNYDTIFEDYEKLLQSENYVTKRQSLKLLGELILDR 239

Qy        241 HNFTIMTKYISKPENLKLMMNLLRDKSRNIQFEAFHVFKVFVANPNKTQPILDILLKNQA 300
              ||| |||||||||||||||||||| ||||||||||||||:|:|||||::||||||
Db        240 HNFAIMTKYISKPENLKLMMNLLRDKSPNIQFEAFHVFKVFVASPHKTQPIVEILLKNQP 299

Qy        301 KLIEFLSKFQNDRTDCMSSSVPTTNSRVDLRVKPRTRGIRDLKRPA 346
              ||||||| || :|||         : :         : |||||: |
Db        300 KLIEFLSSFQKERTD---------DEQFADEKNYLIKQIRDLKKTA 336


RESULT 13
AAQ93064
ID   AAQ93064    PRELIMINARY;      PRT;   337 AA.
AC   AAQ93064;
DT   31-MAR-2004 (TrEMBLrel. 27, Created)
DT   31-MAR-2004 (TrEMBLrel. 27, Last sequence update)
DT   31-MAR-2004 (TrEMBLrel. 27, Last annotation update)
DE   Antigen MLAA-34.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Chen G., Zhang W.G., Fu J., Cao X.M., Chen Y.X., Zhao W.H., Zhao A.Z.,
RA   Han Y.H., Li F.Y., Liu X.P., Yao L.B.;
RT   "Serological identification of immunogenic antigens in acute monocytic
RT   leukemia.";
RL   Submitted (MAY-2003) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AY288977; AAQ93064.1; -.
SQ   SEQUENCE   337 AA;  39029 MW;  B8F1AB5C7905EC93 CRC64;

  Query Match             72.3%;  Score 1287.5;  DB 2;  Length 337;
  Best Local Similarity   75.4%;  Pred. No. 1.8e-74;
  Matches  261;  Conservative   33;  Mismatches   39;  Indels   13;  Gaps    3;

Qy          1 MPFPPFGKSHKSPADIVKNLKESMAVLEKQDISDKKAEKATEEVSKNLVAMKEILYGTNEK 60
              ||  | ||||:||:||| ||:::|:||||   ||| :||:|||||:| |||||| |||:|
Db          4 MPL-FSKSHKNPAEIVKILKDNLAILEKQ---DKKTDKASEEVSKSLQAMKEILCGTNDK 59

Qy         61 EPQTEAVAQLAQELYNSGLLSTLVADLQLIDFEGKKDVAQIFNNILRRQIGTRTPTVEYI 120
              || ||||||||||||:|||| ||:|||||||||||||| ||||||||||||| || |||
Db         60 EPPTEAVAQLAQELYSSGLLVTLIADLQLIDFEGKKDVTQIFNNILRRQIGTRCPTAEYI 119

Qy        121 CTQQNILFMLLKGYESPEIALNCGIMLRECIRHEPLAKIILWSEQFYDFFRYVEMSTFDI 180
                :  :|| ||||||||:|:||| |||||||||||||||||:| || |||:|||:|||||
Db        120 SSHPHILVMLLKGYEAPQIALRCGIMLRECIRHEPLAKIILFSNQFRDFFKYVELSTFDI 179

Qy        181 ASDAFATFKDLLTRHKLLSAEFLEQHYDRFFSEYEKLLHSENYVTKRQSLKLLGELLLDR 240
              |||||||||||||||:| |:||||:||  | :||||| |||||||||||||||:|||
Db        180 ASDAFATFKDLLTRHKVLVADFLEQNYDTIFEDYEKLLQSENYVTKRQSLKLLGELILDR 239

Qy        241 HNFTIMTKYISKPENLKLMMNLLRDKSRNIQFEAFHVFKVFVANPNKTQPILDILLKNQA 300
              ||| |||||||||||||||||||| ||||||||||||||:|:|||||::||||||
Db        240 HNFTIMTKYISKPENLKLMMNLLRDKSPNIQFEAFHVFKVFVASPHKTQPIVEILLKNQP 299

Qy        301 KLIEFLSKFQNDRTDCMSSSVPTTNSRVDLRVKPRTRGIRDLKRPA 346
              ||||||| || :|||         : :         : |||||: |
Db        300 KLIEFLSSFQKERTD---------DEQFADEKNYLIKQIRDLKKAA 336


RESULT 14
Q803V8
ID   Q803V8      PRELIMINARY;      PRT;   343 AA.
AC   Q803V8;
DT   01-JUN-2003 (TrEMBLrel. 24, Created)
DT   01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Zgc:55451.
GN   Name=zgc:55451;
OS   Brachydanio rerio (Zebrafish) (Danio rerio).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes;
OC   Cyprinidae; Danio.
OX   NCBI_TaxID=7955;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=AB; TISSUE=Whole body;
RX   MEDLINE=22388257; PubMed=12477932;
RA   Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA   Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA   Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA   Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA   Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA   Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA   Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA   Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
RA   Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA   Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA   Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA   Fahey J., Helton E., Ketteman M., Madan A., Rodrigues S., Sanchez A.,
RA   Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA   Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA   Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.,
RA   Krzywinski M.I., Skalska U., Smailus D.E., Schnerch A., Schein J.E.,
RA   Jones S.J., Marra M.A.;
RT   "Generation and initial analysis of more than 15,000 full-length human
RT   and mouse cDNA sequences.";
RL   Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=AB; TISSUE=Whole body;
RA   Strausberg R.;
RL   Submitted (JAN-2003) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; BC044172; AAH44172.1; -.
DR   InterPro; IPR008938; ARM.
DR   InterPro; IPR004892; Mo25.
DR   Pfam; PF03204; Mo25; 1.
SQ   SEQUENCE   343 AA;  39820 MW;  F10450DA0446268A CRC64;

  Query Match             72.3%;  Score 1287;  DB 2;  Length 343;
  Best Local Similarity   71.5%;  Pred. No. 2e-74;
  Matches  253;  Conservative   42;  Mismatches   41;  Indels   18;  Gaps    3;

Qy          1 MPFPPFGKSHKSPADIVKNLKESMAVLEKQDISD-KKAEKATEEVSKNLVAMKEILYGTNE 59
              ||||| || ||||:|||:|||::| ||| : |: || ||  ||||||| ::||:| || :
Db          1 MPFPPFEKSQKSPAEIVKSLKENVAYLEKLESSESKKCEKVAEEVSKNLSSLKEVLCGTGD 60

Qy         60 KEPQTEAVAQLAQELYNSGLLSTLVADLQLIDFEGKKDVAQIFNNILRRQIGTRTPTVEY 119
```

```
Db         61 KEPQTEAVAQLAQELYNTNLLISLIANLQRIDFEGKKDVVHLFSNIVRRQIGARTPTVEY 120

Qy        120 ICTQQNILFMLLKGYESPEIALNCGIMLRECIRHEPLAKIILWSEQFYDFFRYVEMSTFD 179

Db        121 ISSHSQILFMLLKGYETSEVALNCGMMLRECLRHDPLARIVLFSEDFYCFFRYVEMSTFD 180

Qy        180 IASDAFATFKDLLTRHKLLSAEFLEQHYDRFFSEYEKLLHSENYVTKRQSLKLLGELLLD 239

Db        181 IASDAFASFKDLLTRHKIMCADFLETNYDRVFTEYEKLLHSENYVTKRQSLKLLGELLLD 240

Qy        240 RHNFTIMTKYISKPENLKLMMNLLRDKSRNIQFEAFHVFKVFVANPNKTQPILDILLKNQ 299

Db        241 RHNFTVATKYISRAENLKLMMNMLRDNSRNIQFEAFHVFKVFVANPNKTQPVLDILLKNQ 300

Qy        300 AKLIEFLSKFQNDRTD----CMSSSVPTTNSRVDLRVKPRTRGIRDLKRPAQQE 349

Db        301 SKLVEFLSHFQTDRSEDEQFCDEKNY-------------LIKQIRDLKRPAPAE 341


RESULT 15
Q7PF07
ID   Q7PF07      PRELIMINARY;      PRT;   332 AA.
AC   Q7PF07;
DT   01-MAR-2004 (TrEMBLrel. 26, Created)
DT   01-MAR-2004 (TrEMBLrel. 26, Last sequence update)
DT   01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE   ENSANGP00000023148 (Fragment).
GN   Name=ENSANGG00000021597;
OS   Anopheles gambiae str. PEST.
OC   Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC   Neoptera; Endopterygota; Diptera; Nematocera; Culicoidea; Anopheles.
OX   NCBI_TaxID=180454;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=PEST;
RA   Anopheles Genome Sequencing Consortium;
RL   Submitted (APR-2003) to the EMBL/GenBank/DDBJ databases.
CC   -!- CAUTION: The sequence shown here is derived from an
CC       EMBL/GenBank/DDBJ whole genome shotgun (WGS) entry which is
CC       preliminary data.
DR   EMBL; AAAB01008804; EAA45510.1; -.
DR   InterPro; IPR004892; Mo25.
DR   Pfam; PF03204; Mo25; 1.
FT   NON_TER     332    332
SQ   SEQUENCE   332 AA;  38665 MW;  E315A6819F3BDA67 CRC64;

  Query Match             65.3%;  Score 1163.5;  DB 2;  Length 332;
  Best Local Similarity   71.5%;  Pred. No. 1.5e-66;
  Matches  226;  Conservative   46;  Mismatches   39;  Indels    5;  Gaps    3;

Qy          1 MPFPPFGKSHKSPADIVKNLKESMAVLEKQDISDKKAEKATEEVSKNLVAMKEILYGTNEK 60

Db          1 MPL-FGKSQKSPQELVKSLKEAVNALER---GDKKAEKAQEDVSKNLVSIKNMLCGTADA 56

Qy         61 EPQTE-AVAQLAQELYNSGLLSTLVADLQLIDFEGKKDVAQIFNNILRRQIGTRTPTVEY 119

Db         57 EPQTEIVISQLAHELYSTDLLLLLIQNLNRIDFEGKKDVAHIFNNVIRRQIGTRLPTVEY 116

Qy        120 ICTQQNILFMLLKGYESPEIALNCGIMLRECIRHEPLAKIILWSEQFYDFFRYVEMSTFD 179

Db        117 ICTKPEILFTLMAGYEHQEIALNCGIMLRECARHEALAKIMLHSEEFFNFFRYVEVSTFD 176

Qy        180 IASDAFATFKDLLTRHKLLSAEFLEQHYDRFFSEYEKLLHSENYVTKRQSLKLLGELLLD 239

Db        177 IASDAFSTFKELLTRHKLLSAEFLEQNYDKVFSRYEALLNSENYVTRRQSLKLLGELLLD 236

Qy        240 RHNFTIMTKYISKPENLKLMMNLLRDKSRNIQFEAFHVFKVFVANPNKTQPILDILLKNQ 299

Db        237 RHNFTVMTKYISNPDNLKLMMNMLKETSRNIQFEAFHVFKVFVANPNKPKPIMDILLRNQ 296

Qy        300 AKLIEFLSKFQNDRTD 315

Db        297 EKLVDFLTRFHTDRSE 312


Search completed: October 22, 2004, 09:12:26
Job time : 197 secs
```

GenCore version 5.1.6

OM protein - protein search, using sw model

Run on: October 22, 2004, 09:12:34 ; Search time 132 Seconds
(without alignments)
858.451 Million cell updates/sec

Title: US-10-089-688-2
Perfect score: 1781
Sequence: 1 MPFPPFGKSHKSPADIVKNLK.....RVKPRTRGIRDLKRPAQQEA 350

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1364641 seqs, 323758627 residues

Total number of hits satisfying chosen parameters: 1364641

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published_Applications_AA:*
1: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep:*
2: /cgn2_6/ptodata/2/pubpaa/PCT_NEW_PUB.pep:*
3: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB.pep:*
4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pep:*
5: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pep:*
6: /cgn2_6/ptodata/2/pubpaa/PCTUS_PUBCOMB.pep:*
7: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pep:*
8: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pep:*
9: /cgn2_6/ptodata/2/pubpaa/US09A_PUBCOMB.pep:*
10: /cgn2_6/ptodata/2/pubpaa/US09B_PUBCOMB.pep:*
11: /cgn2_6/ptodata/2/pubpaa/US09C_PUBCOMB.pep:*
12: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB.pep:*
13: /cgn2_6/ptodata/2/pubpaa/US10A_PUBCOMB.pep:*
14: /cgn2_6/ptodata/2/pubpaa/US10B_PUBCOMB.pep:*
15: /cgn2_6/ptodata/2/pubpaa/US10C_PUBCOMB.pep:*
16: /cgn2_6/ptodata/2/pubpaa/US10D_PUBCOMB.pep:*
17: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB.pep:*
18: /cgn2_6/ptodata/2/pubpaa/US11_NEW_PUB.pep:*
19: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.pep:*
20: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	1635.5	91.8	341	16	US-10-408-765A-911	Sequence 911, App
2	1635.5	91.8	496	14	US-10-239-079-5	Sequence 5, Appli
3	1635.5	91.8	552	14	US-10-239-079-6	Sequence 6, Appli
4	1614.5	90.7	341	14	US-10-025-730-3	Sequence 3, Appli
5	1297.5	72.9	337	14	US-10-025-730-1	Sequence 1, Appli
6	1117	62.7	339	14	US-10-025-730-4	Sequence 4, Appli
7	1054.5	59.2	377	14	US-10-025-730-5	Sequence 5, Appli
8	750	42.1	446	15	US-10-425-114-52177	Sequence 52177, A
9	747	41.9	336	16	US-10-437-963-161075	Sequence 161075,
10	741	41.6	412	15	US-10-425-114-54669	Sequence 54669, A
11	721.5	40.5	339	15	US-10-424-599-224185	Sequence 224185,
12	649	36.4	351	15	US-10-425-114-46025	Sequence 46025, A
13	639.5	35.9	337	16	US-10-437-963-183779	Sequence 183779,
14	635.5	35.7	342	15	US-10-425-114-48872	Sequence 48872, A
15	593	33.3	334	16	US-10-437-963-146612	Sequence 146612,
16	501	28.1	337	15	US-10-424-599-209662	Sequence 209662,
17	400.5	22.5	197	16	US-10-767-701-37391	Sequence 37391, A
18	379.5	21.3	173	15	US-10-424-599-158029	Sequence 158029,
19	353.5	19.8	140	16	US-10-767-701-42587	Sequence 42587, A
20	321	18.0	184	15	US-10-424-599-149595	Sequence 149595,
21	286.5	16.1	294	16	US-10-437-963-119830	Sequence 119830,
22	255	14.3	161	15	US-10-425-114-70029	Sequence 70029, A
23	209	11.7	135	11	US-09-864-408A-6108	Sequence 6108, Ap
24	168	9.4	83	15	US-10-425-114-44450	Sequence 44450, A
25	135.5	7.6	116	15	US-10-424-599-210348	Sequence 210348,
26	126	7.1	1182	15	US-10-282-122A-53445	Sequence 53445, A
27	125.5	7.0	74	15	US-10-425-114-56749	Sequence 56749, A
28	116.5	6.5	888	10	US-09-893-519A-73	Sequence 73, Appl
29	113	6.3	2184	14	US-10-304-095-6	Sequence 6, Appli
30	111.5	6.3	70	15	US-10-424-599-210351	Sequence 210351,
31	111.5	6.3	790	16	US-10-408-765A-2178	Sequence 2178, Ap
32	110.5	6.2	797	16	US-10-437-963-163096	Sequence 163096,
33	109.5	6.1	367	15	US-10-282-122A-45776	Sequence 45776, A
34	109.5	6.1	564	15	US-10-282-122A-61170	Sequence 61170, A
35	109	6.1	807	9	US-09-847-046-2	Sequence 2, Appli
36	109	6.1	825	14	US-10-295-027-546	Sequence 546, App
37	109	6.1	858	14	US-10-367-978-35	Sequence 35, Appl
38	108.5	6.1	719	15	US-10-282-122A-54440	Sequence 54440, A
39	108	6.1	54	16	US-10-437-963-166440	Sequence 166440,
40	105.5	5.9	63	15	US-10-424-599-230655	Sequence 230655,
41	105.5	5.9	660	9	US-09-864-761-47959	Sequence 47959, A
42	105.5	5.9	820	14	US-10-029-386-32324	Sequence 32324, A
43	105.5	5.9	3878	14	US-10-080-608A-11	Sequence 11, Appl
44	105.5	5.9	3899	14	US-10-171-311-4	Sequence 4, Appli
45	105.5	5.9	3907	14	US-10-171-311-2	Sequence 2, Appli

ALIGNMENTS

RESULT 1
US-10-408-765A-911
; Sequence 911, Application US/10408765A
; Publication No. US20040101874A1
; GENERAL INFORMATION:
; APPLICANT: Ghosh, Soumitra S.
; APPLICANT: Fahy, Eoin D.
; APPLICANT: Zhang, Bing
; APPLICANT: Gibson, Bradford W.
; APPLICANT: Taylor, Steven W.
; APPLICANT: Glenn, Gary M.
; APPLICANT: Warnock, Dale E.
; TITLE OF INVENTION: TARGETS FOR THERAPEUTIC INTERVENTION
; TITLE OF INVENTION: IDENTIFIED IN THE MITOCHONDRIAL PROTEOME
; FILE REFERENCE: 660088.465
; CURRENT APPLICATION NUMBER: US/10/408,765A
; CURRENT FILING DATE: 2003-04-04
; NUMBER OF SEQ ID NOS: 3077
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 911
; LENGTH: 341
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-408-765A-911

Query Match 91.8%; Score 1635.5; DB 16; Length 341;
Best Local Similarity 94.0%; Pred. No. 1.5e-129;
Matches 329; Conservative 1; Mismatches 11; Indels 9; Gaps 2;

```
Qy          1 MPFPPFGKSHKSPADIVKNLKESMAVLEKQDISDKKAEKATEEVSKNLVAMKEILYGTNEK 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MPFPPFGKSHKSPADIVKNLKESMAVLEKQDISDKKAEKATEEVSKNLVAMKEILYGTNEK 60

Qy         61 EPQTEAVAQLAQELYNSGLLSTLVADLQLIDFEGKKDVAQIFNNILRRQIGTRTPTVEYI 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 EPQTEAVAQLAQELYNSGLLSTLVADLQLIDFEGKKDVAQIFNNILRRQIGTRTPTVEYI 120
```

```
Qy        121 CTQQNILFMLLKGYESPEIALNCGIMLRECIRHEPLAKIILWSEQFYDFFRYVEMSTFDI 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CTQQNILFMLLKGYESPEIALNCGIMLRECIRHEPLAKIILWSEQFYDFFRYVEMSTFDI 180

Qy        181 ASDAFATFKDLLTRHKLLSAEFLEQHYDRFFSEYEKLLHSENYVTKRQSLKLLGELLLDR 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 ASDAFATFKDLLTRHKLLSAEFLEQHYDRFFSEYEKLLHSENYVTKRQSLKLLGELLLDR 240

Qy        241 HNFTIMTKYISKPENLKLMMNLLRDKSRNIQFEAFHVFKVFVANPNKTQPILDILLKNQA 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 HNFTIMTKYISKPENLKLMMNLLRDKSRNIQFEAFHVFKVFVANPNKTQPILDILLKNQA 300

Qy        301 KLIEFLSKFQNDRTDCMSSSVPTTNSRVDLRVKPRTRGIRDLKRPAQQEA 350
              |||||||||||||:         |       ||      ||||||||||||
Db        301 KLIEFLSKFQNDRTE-----DEQFNDEKTYLVKQ----IRDLKRPAQQEA 341


RESULT 2
US-10-239-079-5
; Sequence 5, Application US/10239079
; Publication No. US20030148446A1
; GENERAL INFORMATION:
;  APPLICANT: Merck Patent GmbH
;  TITLE OF INVENTION: ANIC-BP1-ligand
;  FILE REFERENCE: ANIC-BP-1-ligand
;  CURRENT APPLICATION NUMBER: US/10/239,079
;  CURRENT FILING DATE:  2002-09-19
;  NUMBER OF SEQ ID NOS: 8
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 5
;   LENGTH: 496
;   TYPE: PRT
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Description of Artificial Sequence: Gal4-ANIC-BP-1
;   OTHER INFORMATION: fusion protein
US-10-239-079-5

  Query Match             91.8%;  Score 1635.5;  DB 14;  Length 496;
  Best Local Similarity   94.0%;  Pred. No. 2.5e-129;
  Matches  329;  Conservative    1;  Mismatches   11;  Indels    9;  Gaps    2;

Qy          1 MPFPPFGKSHKSPADIVKNLKESMAVLEKQDISDKKAEKATEEVSKNLVAMKEILYGTNEK 60
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        156 MPFPPFGKSHKSPADIVKNLKESMAVLEKQDISDKKAEKATEEVSKNLVAMKEILYGTNEK 215

Qy         61 EPQTEAVAQLAQELYNSGLLSTLVADLQLIDFEGKKDVAQIFNNILRRQIGTRTPTVEYI 120
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        216 EPQTEAVAQLAQELYNSGLLSTLVADLQLIDFEGKKDVAQIFNNILRRQIGTRTPTVEYI 275

Qy        121 CTQQNILFMLLKGYESPEIALNCGIMLRECIRHEPLAKIILWSEQFYDFFRYVEMSTFDI 180
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        276 CTQQNILFMLLKGYESPEIALNCGIMLRECIRHEPLAKIILWSEQFYDFFRYVEMSTFDI 335

Qy        181 ASDAFATFKDLLTRHKLLSAEFLEQHYDRFFSEYEKLLHSENYVTKRQSLKLLGELLLDR 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        336 ASDAFATFKDLLTRHKLLSAEFLEQHYDRFFSEYEKLLHSENYVTKRQSLKLLGELLLDR 395

Qy        241 HNFTIMTKYISKPENLKLMMNLLRDKSRNIQFEAFHVFKVFVANPNKTQPILDILLKNQA 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        396 HNFTIMTKYISKPENLKLMMNLLRDKSRNIQFEAFHVFKVFVANPNKTQPILDILLKNQA 455

Qy        301 KLIEFLSKFQNDRTDCMSSSVPTTNSRVDLRVKPRTRGIRDLKRPAQQEA 350
              |||||||||||||:         |       ||      ||||||||||||
Db        456 KLIEFLSKFQNDRTE-----DEQFNDEKTYLVKQ----IRDLKRPAQQEA 496


RESULT 3
US-10-239-079-6
; Sequence 6, Application US/10239079
; Publication No. US20030148446A1
; GENERAL INFORMATION:
;  APPLICANT: Merck Patent GmbH
;  TITLE OF INVENTION: ANIC-BP1-ligand
;  FILE REFERENCE: ANIC-BP-1-ligand
;  CURRENT APPLICATION NUMBER: US/10/239,079
;  CURRENT FILING DATE:  2002-09-19
;  NUMBER OF SEQ ID NOS: 8
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 6
;   LENGTH: 552
;   TYPE: PRT
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Description of Artificial Sequence: LexA-ANIC-BP-1
;   OTHER INFORMATION: fusion protein
US-10-239-079-6

  Query Match             91.8%;  Score 1635.5;  DB 14;  Length 552;
  Best Local Similarity   94.0%;  Pred. No. 2.9e-129;
  Matches  329;  Conservative    1;  Mismatches   11;  Indels    9;  Gaps    2;

Qy          1 MPFPPFGKSHKSPADIVKNLKESMAVLEKQDISDKKAEKATEEVSKNLVAMKEILYGTNEK 60
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        212 MPFPPFGKSHKSPADIVKNLKESMAVLEKQDISDKKAEKATEEVSKNLVAMKEILYGTNEK 271

Qy         61 EPQTEAVAQLAQELYNSGLLSTLVADLQLIDFEGKKDVAQIFNNILRRQIGTRTPTVEYI 120
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        272 EPQTEAVAQLAQELYNSGLLSTLVADLQLIDFEGKKDVAQIFNNILRRQIGTRTPTVEYI 331

Qy        121 CTQQNILFMLLKGYESPEIALNCGIMLRECIRHEPLAKIILWSEQFYDFFRYVEMSTFDI 180
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        332 CTQQNILFMLLKGYESPEIALNCGIMLRECIRHEPLAKIILWSEQFYDFFRYVEMSTFDI 391

Qy        181 ASDAFATFKDLLTRHKLLSAEFLEQHYDRFFSEYEKLLHSENYVTKRQSLKLLGELLLDR 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        392 ASDAFATFKDLLTRHKLLSAEFLEQHYDRFFSEYEKLLHSENYVTKRQSLKLLGELLLDR 451

Qy        241 HNFTIMTKYISKPENLKLMMNLLRDKSRNIQFEAFHVFKVFVANPNKTQPILDILLKNQA 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        452 HNFTIMTKYISKPENLKLMMNLLRDKSRNIQFEAFHVFKVFVANPNKTQPILDILLKNQA 511

Qy        301 KLIEFLSKFQNDRTDCMSSSVPTTNSRVDLRVKPRTRGIRDLKRPAQQEA 350
              |||||||||||||:         |       ||      ||||||||||||
Db        512 KLIEFLSKFQNDRTE-----DEQFNDEKTYLVKQ----IRDLKRPAQQEA 552


RESULT 4
US-10-025-730-3
; Sequence 3, Application US/10025730
; Publication No. US20030045466A1
; GENERAL INFORMATION:
;  APPLICANT: Tang, Y. Tom
;  APPLICANT:  Guegler, Karl J.
;  APPLICANT:  Corley, Neil C.
;  APPLICANT:  Gorgone, Gina A.
;  TITLE OF INVENTION: CALCIUM BINDING PROTEIN
;  FILE REFERENCE: PF-0635 US
;  CURRENT APPLICATION NUMBER: US/10/025,730
;  CURRENT FILING DATE:  2001-12-18
;  PRIOR APPLICATION NUMBER: US/09/190,965
;  PRIOR FILING DATE: 1998-11-13
;  NUMBER OF SEQ ID NOS: 5
;  SOFTWARE: PERL Program
; SEQ ID NO 3
;   LENGTH: 341
;   TYPE: PRT
;   ORGANISM: Mus sp.
;   FEATURE: -
;   OTHER INFORMATION: g262934
US-10-025-730-3

  Query Match             90.7%;  Score 1614.5;  DB 14;  Length 341;
  Best Local Similarity   92.9%;  Pred. No. 8.8e-128;
  Matches  325;  Conservative    2;  Mismatches   14;  Indels    9;  Gaps    2;
```

```
Qy          1 MPFPFGKSHKSPADIVKNLKESMAVLEKQDISDKKAEKATEEVSKNLVAMKEILYGTNEK 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MPFPFGKSHKSPADIVKNLKESMAVLEKQDISDKKAEKATEEVSKNLVAMKEILYGTNEK 60

Qy         61 EPQTEAVAQLAQELYNSGLLSTLVADLQLIDFEGKKDVAQIFNNILRRQIGTRTPTVEYI 120
              ||||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||
Db         61 EPQTEAVAQLAQELYNSGLLGTLVADLQLIDFEGKKDVAQIFNNILRRQIGTRTPTVEYI 120

Qy        121 CTQQNILFMLLKGYESPEIALNCGIMLRECIRHEPLAKIILWSEQFYDFFRYVEMSTFDI 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CTQQNILFMLLKGYESPEIALNCGIMLRECIRHEPLAKIILWSEQFYDFFRYVEMSTFDI 180

Qy        181 ASDAFATFKDLLTRHKLLSAEFLEQHYDRFFSEYEKLLHSENYVTKRQSLKLLGELLLDR 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 ASDAFATFKDLLTRHKLLSAEFLEQHYDRFFSEYEKLLHSENYVTKRQSLKLLGELLLDR 240

Qy        241 HNFTIMTKYISKPENLKLMMNLLRDKSRNIQFEAFHVFKVFVANPNKTQPILDILLKNQA 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 HNFTIMTKYISKPENLKLMMNLLRDKSRNIQFEAFHVFKVFVANPNKTQPILDILLKNQT 300

Qy        301 KLIEFLSKFQNDRTDCMSSSVPTTNSRVDLRVKPRTRGIRDLKRPAQQEA 350
              |||||||||||||:        |       ||     ||:||| |||||
Db        301 KLIEFLSKFQNDRTE-----DEQFNDEKTYLVKQ----IRNLKRAAQQEA 341


RESULT 5
US-10-025-730-1
; Sequence 1, Application US/10025730
; Publication No. US20030045466A1
; GENERAL INFORMATION:
;  APPLICANT: Tang, Y. Tom
;  APPLICANT:  Guegler, Karl J.
;  APPLICANT:  Corley, Neil C.
;  APPLICANT:  Gorgone, Gina A.
;  TITLE OF INVENTION: CALCIUM BINDING PROTEIN
;  FILE REFERENCE: PF-0635 US
;  CURRENT APPLICATION NUMBER: US/10/025,730
;  CURRENT FILING DATE:  2001-12-18
;  PRIOR APPLICATION NUMBER: US/09/190,965
;  PRIOR FILING DATE: 1998-11-13
;  NUMBER OF SEQ ID NOS: 5
;  SOFTWARE: PERL Program
; SEQ ID NO 1
;   LENGTH: 337
;   TYPE: PRT
;   ORGANISM: Homo sapiens
;   FEATURE: -
;   OTHER INFORMATION: 3734805
US-10-025-730-1

  Query Match             72.9%;  Score 1297.5;  DB 14;  Length 337;
  Best Local Similarity   76.0%;  Pred. No. 4.7e-101;
  Matches  263;  Conservative   32;  Mismatches   38;  Indels   13;  Gaps    3;

Qy          1 MPFPFGKSHKSPADIVKNLKESMAVLEKQDISDKKAEKATEEVSKNLVAMKEILYGTNEK 60
              ||  | ||||:||:||| ||:::|:|||   ||| :||:|||||:| |||||| |||||
Db          4 MPL-FSKSHKNPAEIVKILKDNLAILEKQ---DKKTDKASEEVSKSLQAMKEILCGTNEK 59

Qy         61 EPQTEAVAQLAQELYNSGLLSTLVADLQLIDFEGKKDVAQIFNNILRRQIGTRTPTVEYI 120
              || |||||||||||:|||| ||:|||||||||||||| ||||||||||||:||||||
Db         60 EPPTEAVAQLAQELYSSGLLVTLIADLQLIDFEGKKDVTQIFNNILRRQIGTRSPTVEYI 119

Qy        121 CTQQNILFMLLKGYESPEIALNCGIMLRECIRHEPLAKIILWSEQFYDFFRYVEMSTFDI 180
                :||||||||:|:||| |||||||||||||||:| || |||:|||:|||||
Db        120 SAHPHILFMLLKGYEAPQIALRCGIMLRECIRHEPLAKIILFSNQFRDFFKYVELSTFDI 179

Qy        181 ASDAFATFKDLLTRHKLLSAEFLEQHYDRFFSEYEKLLHSENYVTKRQSLKLLGELLLDR 240
              ||||||||||||||:| |:||||:||  | :||||| ||||||||||||||||:|||
Db        180 ASDAFATFKDLLTRHKVLVADFLEQNYDTIFEDYEKLLQSENYVTKRQSLKLLGELILDR 239

Qy        241 HNFTIMTKYISKPENLKLMMNLLRDKSRNIQFEAFHVFKVFVANPNKTQPILDILLKNQA 300
              ||| |||||||||||||||||||| ||||||||||||||:|:||||::||||||
Db        240 HNFAIMTKYISKPENLKLMMNLLRDKSPNIQFEAFHVFKVFVASPHKTQPIVEILLKNQP 299

Qy        301 KLIEFLSKFQNDRTDCMSSSVPTTNSRVDLRVKPRTRGIRDLKRPA 346
              ||||||| || :|||         : :         : |||||: |
Db        300 KLIEFLSSFQKERTD---------DEQFADEKNYLIKQIRDLKKTA 336


RESULT 6
US-10-025-730-4
; Sequence 4, Application US/10025730
; Publication No. US20030045466A1
; GENERAL INFORMATION:
;  APPLICANT: Tang, Y. Tom
;  APPLICANT:  Guegler, Karl J.
;  APPLICANT:  Corley, Neil C.
;  APPLICANT:  Gorgone, Gina A.
;  TITLE OF INVENTION: CALCIUM BINDING PROTEIN
;  FILE REFERENCE: PF-0635 US
;  CURRENT APPLICATION NUMBER: US/10/025,730
;  CURRENT FILING DATE:  2001-12-18
;  PRIOR APPLICATION NUMBER: US/09/190,965
;  PRIOR FILING DATE: 1998-11-13
;  NUMBER OF SEQ ID NOS: 5
;  SOFTWARE: PERL Program
; SEQ ID NO 4
;   LENGTH: 339
;   TYPE: PRT
;   ORGANISM: Drosophila melanogaster
;   FEATURE: -
;   OTHER INFORMATION: g1794137
US-10-025-730-4

  Query Match             62.7%;  Score 1117;  DB 14;  Length 339;
  Best Local Similarity   69.0%;  Pred. No. 8e-86;
  Matches  220;  Conservative   51;  Mismatches   40;  Indels    8;  Gaps    5;

Qy          1 MPFPFGKSHKSPADIVKNLKESMAVLEKQDISDKKAEKATEEVSKNLVAMKEILYGTNEK 60
              ||  |||| ||| ::||:|||::  ||     |:| ||| |:||||||::| :|:|:::
Db          1 MPL-FGKSQKSPVELVKSLKEAINALE---AGDRKVEKAQEDVSKNLVSIKNMLHGSSDA 56

Qy         61 EPQTE-AVAQLAQELYNSGLLSTLVADLQLIDFEGKKDVAQIFNNILRRQIGTRTPTVEY 119
              ||  :  ||||:|||||| ||  |: :|  ||||||| || ||||:|||||||:|||||
Db         57 EPPADYVVAQLSQELYNSNLLLLLIQNLHRIDFEGKKHVALIFNNLLRRQIGTRSPTVEY 116

Qy        120 ICTQQNILFMLLKGYES--PEIALNCGIMLRECIRHEPLAKIILWSEQFYDFFRYVEMST 177
              |||:  ||| |: |||   |||||| | ||||| |:| ||||:| |::|: ||||||:||
Db        117 ICTKPEILFTLMAGYEDAHPEIALNSGTMLRECARYEALAKIMLHSDEFFKFFRYVEVST 176

Qy        178 FDIASDAFATFKDLLTRHKLLSAEFLEQHYDRFFSE-YEKLLHSENYVTKRQSLKLLGEL 236
              ||||||||:|||:|||||||| |||: :|:|||: |::||:|||||:|||||||:|||||
Db        177 FDIASDAFSTFKELLTRHKLLCAEFLDANYDKFFSQHYQRLLNSENYVTRRQSLKLLGEL 236

Qy        237 LLDRHNFTIMTKYISKPENLKLMMNLLRDKSRNIQFEAFHVFKVFVANPNKTQPILDILL 296
              ||||||||:||:|||:|||||||||:|::|||||||||||||||||||| :|||||||
Db        237 LLDRHNFTVMTRYISEPENLKLMMNMLKEKSRNIQFEAFHVFKVFVANPNKPKPILDILL 296

Qy        297 KNQAKLIEFLSKFQNDRTD 315
              :|| ||::||: |  ||::
Db        297 RNQTKLVDFLTNFHTDRSE 315


RESULT 7
US-10-025-730-5
; Sequence 5, Application US/10025730
; Publication No. US20030045466A1
; GENERAL INFORMATION:
;  APPLICANT: Tang, Y. Tom
;  APPLICANT:  Guegler, Karl J.
;  APPLICANT:  Corley, Neil C.
;  APPLICANT:  Gorgone, Gina A.
;  TITLE OF INVENTION: CALCIUM BINDING PROTEIN
;  FILE REFERENCE: PF-0635 US
;  CURRENT APPLICATION NUMBER: US/10/025,730
```

```
;  CURRENT FILING DATE:  2001-12-18
;  PRIOR APPLICATION NUMBER: US/09/190,965
;  PRIOR FILING DATE: 1998-11-13
;  NUMBER OF SEQ ID NOS: 5
;  SOFTWARE: PERL Program
; SEQ ID NO 5
;   LENGTH: 377
;   TYPE: PRT
;   ORGANISM: Caenorhabditis elegans
;   FEATURE: -
;   OTHER INFORMATION: g1255838
US-10-025-730-5

  Query Match             59.2%;  Score 1054.5;  DB 14;  Length 377;
  Best Local Similarity   62.5%;  Pred. No. 1.7e-80;
  Matches  205;  Conservative   49;  Mismatches   61;  Indels   13;  Gaps    2;

Qy          1 MPFPPGKSHKSPADIVKNLKESMAVLEK----------QDISDKKAEKATEEVSKNLVAM 50
              ||  ||||||||||:|| |:| : :|:|              |||| :|| :|||||:  :
Db          1 MPLLFGKSHKSPADVVKTLREVLTILDKLPPPKLDKDGNIQSDKKYDKALDEVSKNVAMI 60

Qy         51 KEILYGTNEKEPQTE---AVAQLAQELYNSGLLSTLVADLQLIDFEGKKDVAQIFNNILR 107
              |  :|| :  || :|    |||||||:||: :|  |:  |   :|| |||| |||||:||
Db         61 KSFIYGNDSAEPSSEHVVQVAQLAQEVYNANILPMLIKMLPKFEFECKKDVGQIFNNLLR 120

Qy        108 RQIGTRTPTVEYICTQQNILFMLLKGYESPEIALNCGIMLRECIRHEPLAKIILWSEQFY 167
              ||||||:|||||:  :  ||  |::||  |:||| ||:|||| |||: ||||||:|: ||
Db        121 RQIGTRSPTVEYLGARPEILIQLVQGYSVPDIALTCGLMLRESIRHDHLAKIILYSDVFY 180

Qy        168 DFFRYVEMSTFDIASDAFATFKDLLTRHKLLSAEFLEQHYDRFFSEYEKLLHSENYVTKR 227
              || ||:   |||:||||:|||:| |||| : |||| :|| ||::|: ||:|:||||:|
Db        181 TFFLYVQSEVFDISSDAFSTFKELTTRHKAIIAEFLDSNYDTFFAQYQNLLNSKNYVTRR 240

Qy        228 QSLKLLGELLLDRHNFTIMTKYISKPENLKLMMNLLRDKSRNIQFEAFHVFKVFVANPNK 287
              |||||||||||||||  |||||| |:||:||| ||||||||||:|||||||||||||||
Db        241 QSLKLLGELLLDRHNFNTMTKYISNPDNLRLMMELLRDKSRNIQYEAFHVFKVFVANPNK 300

Qy        288 TQPILDILLKNQAKLIEFLSKFQNDRTD 315
               :|| ||| :|: ||:||||:| |||||
Db        301 PKPISDILNRNREKLVEFLSEFHNDRTD 328


RESULT 8
US-10-425-114-52177
; Sequence 52177, Application US/10425114
; Publication No. US20040034888A1
; GENERAL INFORMATION:
;  APPLICANT: Liu, Jingdong
;  APPLICANT:  Zhou, Yihua
;  APPLICANT:  Kovalic, David K.
;  APPLICANT:  Screen, Steven E
;  APPLICANT:  Tabaska, Jack E
;  APPLICANT:  Cao, Yongwei
;  TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
;  TITLE OF INVENTION:  Plants and Uses Thereof for Plant Improvement
;  FILE REFERENCE: 38-21(53313)B
;  CURRENT APPLICATION NUMBER: US/10/425,114
;  CURRENT FILING DATE:  2003-04-28
;  NUMBER OF SEQ ID NOS: 73128
; SEQ ID NO 52177
;   LENGTH: 446
;   TYPE: PRT
;   ORGANISM: Zea mays
;   FEATURE:
;   OTHER INFORMATION: Clone ID: LIB3912-007-H10_FLI.pep
US-10-425-114-52177

  Query Match             42.1%;  Score 750;  DB 15;  Length 446;
  Best Local Similarity   46.7%;  Pred. No. 1e-54;
  Matches  147;  Conservative   64;  Mismatches  100;  Indels    4;  Gaps    2;

Qy          5 FGKSHKSPADIVKNLKESMAVLEKQDIS---DKKAEKATEEVSKNLVAMKEILYGTNEKE 61
              |    ::| |||:  :| :  |:    |   | | |:   |:|||:  || ||||  | |
Db        115 FKSKPRTPVDIVRQTRECLVYLDLHSDSRSGDAKREEKMTELSKNIRDMKSILYGNGESE 174

Qy         62 PQTEAVAQLAQELYNSGLLSTLVADLQLIDFEGKKDVAQIFNNILRRQIGTRTPTVEYIC 121
              | |||  || || :    |  |:  |  :: | :||  |:  |: |:|: ::    ||:
Db        175 PVTEACVQLTQEFFRENTLRLLIIHLPKLNLETRKDATQVVANLQRQQVSSKIVASEYLE 234

Qy        122 TQQNILFMLLKGYESPEIALNCGIMLRECIRHEPLAKIILWSEQFYDFFRYVEMSTFDIA 181
              : :::| :|: |||: :|||: | ||||||||: :|: :| ||    || |:::  ||||
Db        235 SNKDLLDILILGYENMDIALHYGAMLRECIRHQSIARYVLESEHMKKFFDYIQLPNFDIA 294

Qy        182 SDAFATFKDLLTRHKLLSAEFLEQHYDRFFSEY-EKLLHSENYVTKRQSLKLLGELLLDR 240
              ||| ||||:|||||   ||||  :|| || |:  :|| | ||:||||::||||::||||
Db        295 SDASATFKELLTRHKATVAEFLSNNYDWFFEEFNSRLLSSTNYITKRQAIKLLGDMLLDR 354

Qy        241 HNFTIMTKYISKPENLKLMMNLLRDKSRNIQFEAFHVFKVFVANPNKTQPILDILLKNQA 300
              |  :| :|:|  :|| ::|||||| |:||| |||||||:| || ||   :::||: |::
Db        355 SNAAVMMRYVSSKDNLMILMNLLRDSSKNIQIEAFHVFKLFAANKNKPPEVVNILVTNRS 414

Qy        301 KLIEFLSKFQNDRTD 315
              ||: | : |: |: |
Db        415 KLLRFFAGFKIDKED 429


RESULT 9
US-10-437-963-161075
; Sequence 161075, Application US/10437963
; Publication No. US20040123343A1
; GENERAL INFORMATION:
;  APPLICANT: La Rosa, Thomas J.
;  APPLICANT:  Kovalic, David K.
;  APPLICANT:  Zhou, Yihua
;  APPLICANT:  Cao, Yongwei
;  APPLICANT:  Wu, Wei
;  APPLICANT:  Boukharov, Andrey A.
;  APPLICANT:  Barbazuk, Brad
;  APPLICANT:  Li, Ping
;  TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With
;  TITLE OF INVENTION:  Plants and Uses Thereof for Plant Improvement
;  FILE REFERENCE: 38-21(53221)B
;  CURRENT APPLICATION NUMBER: US/10/437,963
;  CURRENT FILING DATE:  2003-05-14
;  NUMBER OF SEQ ID NOS: 204966
; SEQ ID NO 161075
;   LENGTH: 336
;   TYPE: PRT
;   ORGANISM: Oryza sativa
;   FEATURE:
;   OTHER INFORMATION: Clone ID: PAT_MRT4530_60293C.1.pep
US-10-437-963-161075

  Query Match             41.9%;  Score 747;  DB 16;  Length 336;
  Best Local Similarity   46.3%;  Pred. No. 1.3e-54;
  Matches  146;  Conservative   66;  Mismatches   99;  Indels    4;  Gaps    2;

Qy          5 FGKSHKSPADIVKNLKESMAVLEKQDIS---DKKAEKATEEVSKNLVAMKEILYGTNEKE 61
              |    ::|||:|:  :| :  |:    |   | | |:   |:|||:  :| ||||  | |
Db          5 FKSKPRTPADVVRQTRELLIFLDLHSGSRGGDAKREEKMAELSKNIRELKSILYGNGESE 64

Qy         62 PQTEAVAQLAQELYNSGLLSTLVADLQLIDFEGKKDVAQIFNNILRRQIGTRTPTVEYIC 121
              | |||  || || :    |  |:  |  :: | :||  |:  |: |:|: ::    ||:
Db         65 PVTEACVQLTQEFFRENTLRLLIICLPKLNLETRKDATQVVANLQRQQVSSKIVASEYLE 124

Qy        122 TQQNILFMLLKGYESPEIALNCGIMLRECIRHEPLAKIILWSEQFYDFFRYVEMSTFDIA 181
              :::|  |: |||: :|||: | ||||||||: :|: :| |:    || |:::  ||||
Db        125 ANKDLLDTLISGYENMDIALHYGSMLRECIRHQSIARYVLESDHMKKFFDYIQLPNFDIA 184

Qy        182 SDAFATFKDLLTRHKLLSAEFLEQHYDRFFSEYE-KLLHSENYVTKRQSLKLLGELLLDR 240
              ||| ||||:|||||   |||| ::|| ||||:  :|| | ||:||||::| ||::||||
Db        185 SDASATFKELLTRHKATVAEFLSKNYDWFFSEFNTRLLSSTNYITKRQAIKFLGDMLLDR 244

Qy        241 HNFTIMTKYISKPENLKLMMNLLRDKSRNIQFEAFHVFKVFVANPNKTQPILDILLKNQA 300
              | |:| :|:|  :|| ::|||||| |:||| |||||||:| || ||   :::||: |::
Db        245 SNSTVMMRYVSSKDNLMILMNLLRDSSKNIQIEAFHVFKLFAANKNKPTEVVNILVTNRS 304
```

```
Qy        301 KLIEFLSKFQNDRTD 315
              ||: | : |: |: |
Db        305 KLLRFFAGFKIDKED 319


RESULT 10
US-10-425-114-54669
; Sequence 54669, Application US/10425114
; Publication No. US20040034888A1
; GENERAL INFORMATION:
;  APPLICANT: Liu, Jingdong
;  APPLICANT:  Zhou, Yihua
;  APPLICANT:  Kovalic, David K.
;  APPLICANT:  Screen, Steven E
;  APPLICANT:  Tabaska, Jack E
;  APPLICANT:  Cao, Yongwei
;  TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
;  TITLE OF INVENTION:  Plants and Uses Thereof for Plant Improvement
;  FILE REFERENCE: 38-21(53313)B
;  CURRENT APPLICATION NUMBER: US/10/425,114
;  CURRENT FILING DATE:  2003-04-28
;  NUMBER OF SEQ ID NOS: 73128
; SEQ ID NO 54669
;   LENGTH: 412
;   TYPE: PRT
;   ORGANISM: Zea mays
;   FEATURE:
;   OTHER INFORMATION: Clone ID: LIB3069-003-D7_FLI.pep
US-10-425-114-54669

  Query Match             41.6%;  Score 741;  DB 15;  Length 412;
  Best Local Similarity   46.3%;  Pred. No. 5.4e-54;
  Matches  146;  Conservative   64;  Mismatches  101;  Indels    4;  Gaps    2;

Qy          5 FGKSHKSPADIVKNLKESMAVLEKQDIS---DKKAEKATEEVSKNLVAMKEILYGTNEKE 61
              |    ::| |||:  :| :  |:    |   | | ::   |:|||:  || ||||  | |
Db         81 FKTKPRTPVDIVRQTRECLVHLDLHSGSRSGDAKRDEKMTELSKNIRDMKSILYGNGESE 140

Qy         62 PQTEAVAQLAQELYNSGLLSTLVADLQLIDFEGKKDVAQIFNNILRRQIGTRTPTVEYIC 121
              | |||  || || :    | |:  |  :: | :||  |:  |: |:|: ::    ||:
Db        141 PVTEACVQLTQEFFRENTLRLLIIHLPKLNLETRKDATQVVANLQRQQVSSKIVASEYLE 200

Qy        122 TQQNILFMLLKGYESPEIALNCGIMLRECIRHEPLAKIILWSEQFYDFFRYVEMSTFDIA 181
              : :::| :|: |||: :|||: | |||||||: :|: :| ||    || |:::  ||||
Db        201 SNKDLLDILILGYENMDIALHYGAMLRECIRHQSIARYVLESEHMKKFFDYIQLPNFDIA 260

Qy        182 SDAFATFKDLLTRHKLLSAEFLEQHYDRFFSEY-EKLLHSENYVTKRQSLKLLGELLLDR 240
              ||| ||||:|||||   ||||  :|| || |:  :|| | ||:||||::||||::||||
Db        261 SDASATFKELLTRHKATVAEFLSNNYDWFFEEFNSRLLSSTNYITKRQAIKLLGDMLLDR 320

Qy        241 HNFTIMTKYISKPENLKLMMNLLRDKSRNIQFEAFHVFKVFVANPNKTQPILDILLKNQA 300
              |  :| :|:|  :|| ::|||||| |:||| |||||||:| || ||   :::||: |:
Db        321 SNVAVMMRYVSSKDNLMILMNLLRDSSKNIQIEAFHVFKLFAANKNKPPEVVNILVTNRN 380

Qy        301 KLIEFLSKFQNDRTD 315
              ||: | : |: |: |
Db        381 KLLRFFAGFKIDKED 395


RESULT 11
US-10-424-599-224185
; Sequence 224185, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:
;  APPLICANT: La Rosa Thomas J
;  APPLICANT:  Kovalic David K
;  APPLICANT:  Zhou Yihua
;  APPLICANT:  Cao Yongwei
;  TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
;  TITLE OF INVENTION:  Plants and Uses Thereof for Plant Improvement
;  FILE REFERENCE: 38-21(53223)B
;  CURRENT APPLICATION NUMBER: US/10/424,599
;  CURRENT FILING DATE:  2003-04-28
;  NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 224185
;   LENGTH: 339
;   TYPE: PRT
;   ORGANISM: Glycine max
;   FEATURE:
;   OTHER INFORMATION: Clone ID: PAT_MRT3847_44468C.1.pep
US-10-424-599-224185

  Query Match             40.5%;  Score 721.5;  DB 15;  Length 339;
  Best Local Similarity   48.5%;  Pred. No. 1.8e-52;
  Matches  150;  Conservative   58;  Mismatches   98;  Indels    3;  Gaps    2;

Qy          1 MPFPFGKS--HKSPADIVKNLKESMAVLEKQDISDKKA-EKATEEVSKNLVAMKEILYGT 57
              | | | |:   |:| :: |::|||:  |: : : : || ||| ||| || | |: :| |
Db          1 MSFSFFKALRPKTPQEVAKSIKESLMALDTKTVVEVKALEKALEEVEKNFVTMRTMLSGD 60

Qy         58 NEKEPQTEAVAQLAQELYNSGLLSTLVADLQLIDFEGKKDVAQIFNNILRRQIGTRTPTV 117
               | ||  : |:|| :|:    :|: |:  | :: :| :||:   :: :|: :: |    |
Db         61 GESEPNLDQVSQLVEEICKEDVLTLLIHKLPILGWEARKDLVHCWSILLKHKVETNYYCV 120

Qy        118 EYICTQQNILFMLLKGYESPEIALNCGIMLRECIRHEPLAKIILWSEQFYDFFRYVEMST 177
              |||      :|  |:  |:: :|||:||||||||:   ||: || |  |  ||::||:
Db        121 EYIEQHIELLDFLVVCYDNKDIALSCGIMLRECIKFPSLARYILESASFVLFFKFVELPN 180

Qy        178 FDIASDAFATFKDLLTRHKLLSAEFLEQHYDRFFSEYEKLLHSENYVTKRQSLKLLGELL 237
              ||:|||||:|||||||:|  : :|||  ||| ||  |||||  | ||||:|||||||  | |
Db        181 FDVASDAFSTFKDLLTKHVNVVSEFLTAHYDEFFDLYEKLLTSPNYVTRRQSLKLLSEFL 240

Qy        238 LDRHNFTIMTKYISKPENLKLMMNLLRDKSRNIQFEAFHVFKVFVANPNKTQPILDILLK 297
              |:  |  || :|| :   ||:|| |||| |:|||  |||:|||||||||| : :  || |
Db        241 LESPNSQIMKQYILEVRYLKVMMTLLRDSSKNIQLSAFHIFKVFVANPNKPREVKIILSK 300

Qy        298 NQAKLIEFL 306
              || ||:: |
Db        301 NQEKLLDLL 309


RESULT 12
US-10-425-114-46025
; Sequence 46025, Application US/10425114
; Publication No. US20040034888A1
; GENERAL INFORMATION:
;  APPLICANT: Liu, Jingdong
;  APPLICANT:  Zhou, Yihua
;  APPLICANT:  Kovalic, David K.
;  APPLICANT:  Screen, Steven E
;  APPLICANT:  Tabaska, Jack E
;  APPLICANT:  Cao, Yongwei
;  TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
;  TITLE OF INVENTION:  Plants and Uses Thereof for Plant Improvement
;  FILE REFERENCE: 38-21(53313)B
;  CURRENT APPLICATION NUMBER: US/10/425,114
;  CURRENT FILING DATE:  2003-04-28
;  NUMBER OF SEQ ID NOS: 73128
; SEQ ID NO 46025
;   LENGTH: 351
;   TYPE: PRT
;   ORGANISM: Zea mays
;   FEATURE:
;   OTHER INFORMATION: Clone ID: 700170954_FLI.pep
US-10-425-114-46025

  Query Match             36.4%;  Score 649;  DB 15;  Length 351;
  Best Local Similarity   44.0%;  Pred. No. 2.5e-46;
  Matches  136;  Conservative   62;  Mismatches  101;  Indels   10;  Gaps    2;

Qy         12 PADIVKNLKESMAVLEKQDISDKKAEKATEEVSKNLVAMKEILYGTNEKEPQTEAVAQLA 71
              | :| ::||:|:  |:       |   || |:  ||:: ::  | |  | ||  | | |:|
Db         30 PQEIARSLKDSLVALD-----TKTGAKALEDAEKNILTLRHTLAGDGEVEPNQEQVLQIA 84

Qy         72 QELYNSGLLSTLVADLQLIDFEGKKDVAQIFNNILRRQIGTRTPTVEYICTQQNILFMLL 131
```

```
                 |:   |:||  | :|  : :| :||:   :  :||:::      |:||    ::|  |:
Db         85 LEICKEGVLSLFVQNLPSLGWEARKDLVHCWCILLRQKVDESYCCVQYIENHFDLLDFLV 144

Qy        132 KGYESPEIALNCGIMLRECIRHEPLAKIILWSEQFYDFFRYVEMSTFDIASDAFATFKDL 191
              |:: |:|||| ||||||::  ||| || |  |  ||:|||:| |||||||  |||||
Db         145 VCYKNLEVALNCGNMLRECIKYPTLAKYILESSSFELFFQYVELSNFDIASDALNTFKDL 204

Qy        192 LTRHKLLSAEFLEQHYDRFFSEYEKLLHSENYVTKRQSLKLLGELLLDRHNFTIMTKYIS 251
              ||:|:   :|||  ||::||  | ||| | ||||:|||:| | | ||:  |  || :||
Db         205 LTKHEDAVSEFLSSHYEQFFGLYTKLLSSTNYVTRRQSVKFLSEFLLEAPNAQIMKRYIL 264

Qy        252 KPENLKLMMNLLRDKSRNIQFEAFHVFKVFVANPNKTQPILDILLKNQAKLIEFL----- 306
              :  | :|| ||:| |:||:  |||:|||||||||| : |:  |::|  :::: |
Db         265 EVHYLNIMMGLLKDSSKNIRICAFHIFKVFVANPNKPREIIQFLVENHREVLKLLHNLPT 324

Qy        307 SKFQNDRTD 315
              || :::: |
Db         325 SKGEDEQLD 333


RESULT 13
US-10-437-963-183779
; Sequence 183779, Application US/10437963
; Publication No. US20040123343A1
; GENERAL INFORMATION:
;  APPLICANT: La Rosa, Thomas J.
;  APPLICANT:  Kovalic, David K.
;  APPLICANT:  Zhou, Yihua
;  APPLICANT:  Cao, Yongwei
;  APPLICANT:  Wu, Wei
;  APPLICANT:  Boukharov, Andrey A.
;  APPLICANT:  Barbazuk, Brad
;  APPLICANT:  Li, Ping
;  TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With
;  TITLE OF INVENTION:  Plants and Uses Thereof for Plant Improvement
;  FILE REFERENCE: 38-21(53221)B
;  CURRENT APPLICATION NUMBER: US/10/437,963
;  CURRENT FILING DATE:  2003-05-14
;  NUMBER OF SEQ ID NOS: 204966
; SEQ ID NO 183779
;   LENGTH: 337
;   TYPE: PRT
;   ORGANISM: Oryza sativa
;   FEATURE:
;   NAME/KEY: unsure
;   LOCATION: (1)..(337)
;   OTHER INFORMATION: unsure at all Xaa locations
;   FEATURE:
;   OTHER INFORMATION: Clone ID: PAT_MRT4530_80837C.1.pep
US-10-437-963-183779

  Query Match             35.9%;  Score 639.5;  DB 16;  Length 337;
  Best Local Similarity   40.4%;  Pred. No. 1.5e-45;
  Matches  135;  Conservative   73;  Mismatches  109;  Indels   17;  Gaps    2;

Qy         11 SPADIVKNLKESMAVLEKQDISDKKAEKATEEVSKNLVAMKEILYGTNEKEPQTEAVAQL 70
              || ::|:::|||:  |:      :   || |:| ||:  ::: | |  | ||  | | |:
Db         15 SPQELVRSIKESLLALD-----TRTGAKALEDVEKNVSTLRQTLSGDGEVEPNQEQVLQI 69

Qy         71 AQELYNSGLLSTLVADLQLIDFEGKKDVAQIFNNILRRQIGTRTPTVEYICTQQNILFML 130
              | |:    :||  | ::  : :||:||:|  :: :||:::      |:||    ::|  |
Db         70 ALEICKEDVLSLFVQNMPSLGWEGRKDLAHCWSILLRQKVDEAYCCVQYIENHFDLLDFL 129

Qy        131 LKGYESPEIALNCGIMLRECIRHEPLAKIILWSEQFYDFFRYVEMSTFDIASDAFATFKD 190
              :  |:: |:|||| ||||||::  ||| || |  |  ||:|||:| |||||||  ||||
Db         130 VVCYKNLEVALNCGNMLRECIKYPTLAKYILESSSFELFFQYVELSNFDIASDALNTFKD 189

Qy        191 LLTRHKLLSAEFLEQHYDRFFSEYEKLLHSENYVTKRQSLKLLGELLLDRHNFTIMTKYI 250
              |||:|:   :|||  ||::||  | :|| | ||||:|||:| | | ||:  |  || :||
Db         190 LLTKHEAAVSEFLCSHYEQFFELYTRLLTSTNYVTRRQSVKFLSEFLLEAPNAQIMKRYI 249

Qy        251 SKPENLKLMMNLLRDKSRNIQFEAFHVFKVFVANPNKTQPILDILLKNQAKLIEFLSKFQ 310
              :   | :|: ||:   :||:  |||:|||||||||| : |: :|: |  :|:: |
Db         250 VEVSYLNIMIGLLKFSXKNIRICAFHIFKVFVANPNKPRDIIQVLVDNHRELLKLL---- 305

Qy        311 NDRTDCMSSSVPTTNSRVDLRVKPRTRGIRDLKR 344
                      ::||:    :   : }   |:::::
Db         306 --------GNLPTSKGEDEQLEEERDLIIKEIEK 331


RESULT 14
US-10-425-114-48872
; Sequence 48872, Application US/10425114
; Publication No. US20040034888A1
; GENERAL INFORMATION:
;  APPLICANT: Liu, Jingdong
;  APPLICANT:  Zhou, Yihua
;  APPLICANT:  Kovalic, David K.
;  APPLICANT:  Screen, Steven E
;  APPLICANT:  Tabaska, Jack E
;  APPLICANT:  Cao, Yongwei
;  TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
;  TITLE OF INVENTION:  Plants and Uses Thereof for Plant Improvement
;  FILE REFERENCE: 38-21(53313)B
;  CURRENT APPLICATION NUMBER: US/10/425,114
;  CURRENT FILING DATE:  2003-04-28
;  NUMBER OF SEQ ID NOS: 73128
; SEQ ID NO 48872
;   LENGTH: 342
;   TYPE: PRT
;   ORGANISM: Zea mays
;   FEATURE:
;   OTHER INFORMATION: Clone ID: LIB3637-244-E1_FLI.pep
US-10-425-114-48872

  Query Match             35.7%;  Score 635.5;  DB 15;  Length 342;
  Best Local Similarity   43.0%;  Pred. No. 3.3e-45;
  Matches  131;  Conservative   66;  Mismatches  103;  Indels    5;  Gaps    1;

Qy         11 SPADIVKNLKESMAVLEKQDISDKKAEKATEEVSKNLVAMKEILYGTNEKEPQTEAVAQL 70
              :| ::|:::|:|    |        :   || ||| ||: ::: :::|  | ||  | | |:
Db         25 TPEEVVRSIKDSFLALH-----TRTHAKALEEVEKNMSSLRLLIFGDGEVEPNEEQVLQI 79

Qy         71 AQELYNSGLLSTLVADLQLIDFEGKKDVAQIFNNILRRQIGTRTPTVEYICTQQNILFML 130
              |:    ::| :| ||  : :  :||:|  :  :||:::      |:|:     :|  |
Db         80 TLEICKEDVISLIVQDLPSLGWGVRKDLAICWCILLRQKVDETYCCVQYLENHLELLDFL 139

Qy        131 LKGYESPEIALNCGIMLRECIRHEPLAKIILWSEQFYDFFRYVEMSTFDIASDAFATFKD 190
              :  |:: :|||||| ||||||::  ||| || |  |  || |||:  |||||||  ||||
Db         140 VGCYKNLDIALNCGNMLRECIKYPTLAKYILESGSFELFFEYVELPNFDIASDALNTFKD 199

Qy        191 LLTRHKLLSAEFLEQHYDRFFSEYEKLLHSENYVTKRQSLKLLGELLLDRHNFTIMTKYI 250
              |||:|::: ||||  ||::||  | :|| | ||||:||::| | | ||: ||  || :||
Db         200 LLTKHEVVVAEFLSSHYEQFFELYSRLLSSTNYVTRRQAIKFLSEFLLETHNSQIMKRYI 259

Qy        251 SKPENLKLMMNLLRDKSRNIQFEAFHVFKVFVANPNKTQPILDILLKNQAKLIEFLSKFQ 310
              :   | :|:|||:| |:||:  |||||||||||||| : |:  || |: :::: |
Db         260 VEVRFLNIMINLLKDSSKNIRICAFHVFKVFVANPNKPRCIVVALLDNRREVLKLLHNLP 319

Qy        311 NDRTD 315
              : : |
Db         320 SSKGD 324


RESULT 15
US-10-437-963-146612
; Sequence 146612, Application US/10437963
; Publication No. US20040123343A1
; GENERAL INFORMATION:
;  APPLICANT: La Rosa, Thomas J.
;  APPLICANT:  Kovalic, David K.
;  APPLICANT:  Zhou, Yihua
;  APPLICANT:  Cao, Yongwei
;  APPLICANT:  Wu, Wei
;  APPLICANT:  Boukharov, Andrey A.
;  APPLICANT:  Barbazuk, Brad
```

; APPLICANT: Li, Ping
; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53221)B
; CURRENT APPLICATION NUMBER: US/10/437,963
; CURRENT FILING DATE: 2003-05-14
; NUMBER OF SEQ ID NOS: 204966
; SEQ ID NO 146612
; LENGTH: 334
; TYPE: PRT
; ORGANISM: Oryza sativa
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT4530_4721C.1.pep
US-10-437-963-146612

Query Match 33.3%; Score 593; DB 16; Length 334;
Best Local Similarity 39.2%; Pred. No. 1.2e-41;
Matches 125; Conservative 65; Mismatches 97; Indels 32; Gaps 4;

```
Qy          5 FGKSHKSPADIVKNLKESMAVLEKQ-----DISDKKAEKATEEVSKNLVAMKEILYGTNE 59
              |    : | ::|:: :| :| | :       |  | | |    :::|::  || ||||  |
Db         23 FRSKARGPVEVVRHARELLAFLAENHDACGDKRDVKREHKMVDLAKSIGEMKSILYGNGE 82

Qy         60 KEPQTEAVAQLAQELY--NSGLLSTLVADLQLIDFEGKKDVAQIFNNILRRQIGTRTPTV 117
               :|  || :|| :| :  |:  |  ||  |  :| | :||| |:  |:|
Db         83 ADPVDEACSQLTKEFFKENTNSLHLLVVCLPYMDLETQKDVTQVTANLLG---------- 132

Qy        118 EYICTQQNILFMLLKGYESPEIALNCGIMLRECIRHEPLAKIILWSEQFYDFFRYVEMST 177
                        | |:: :||::   :||:||||:  |: :| |:    || |::
Db        133 -------------KRYDNMDIAIHYSAILRDCIRHQVAARYVLESQHMKKFFDYIQYPD 178

Qy        178 FDIASDAFATFKDLLTRHKLLSAEFLEQHYDRFFSEY-EKLLHSENYVTKRQSLKLLGEL 236
              |:|||||| |||:|||||:  :|||   :|| || |:  ||| | ||: :||: :|||::
Db        179 FNIASDAFKTFKELLTRHRSSAAEFFSNNYDWFFPEFNSKLLSSSNYIIRRQATQLLGDI 238

Qy        237 LLDRHNFTIMTKYISKPENLKLMMNLLRDKSRNIQFEAFHVFKVFVANPNKTQPILDILL 296
              |||: | | | :|:|  :|| ::||||||::|| || ||| |||:|  || |   |: ||:
Db        239 LLDKSNTTAMVRYVSSKDNLIILMNLLREQSRAIQVEAFRVFKLFTCNPKKPPEIVGILV 298

Qy        297 KNQAKLIEFLSKFQNDRTD 315
              |::|:: ||: |  :: |
Db        299 TNKSKILRFLADFTIEKED 317
```

Search completed: October 22, 2004, 09:24:17
Job time : 133 secs